PARAMETERSCHÄTZUNG FÜR HILBERTRAUMWERTIGE STOCHASTISCHE PROZESSE

DISSERTATION

zur Erlangung des Grades eines
Doktors der Naturwissenschaften
der Abteilung für Mathematik
an der Ruhr-Universität Bochum

von

WILFRIED LOGES

Bochum 1982

Digitized by the Internet Archive
in 2026

https://archive.org/details/IA41549226_0006

PARAMETERSCHÄTZUNG FÜR HILBERTRAUMWERTIGE STOCHASTISCHE PROZESSE

DISSERTATION

zur Erlangung des Grades eines
Doktors der Naturwissenschaften
der Abteilung für Mathematik
an der Ruhr-Universität Bochum

von

WILFRIED LOGES

Bochum 1982

UNIV.
LEIDEN
BIBL.

Dissertation eingereicht am: 11.10.1982

Tag der mündlichen Prüfung : 1.12.1982

Dekan : Prof. Dr. D. Braess

Referent : Prof. Dr. S. Albeverio

Korreferent: Prof. Dr. V. Baumann

Inhaltsverzeichnis

Kapitel I: Einleitung

Kapitel II: Parameterschätzung für autoregressive Prozesse bei indirekter Beobachtung

§ II.1 Erläuterung des Modells

§ II.2 Herleitung eines modifizierten Kleinste-Quadrate-Schätzers für die Systemparameter des autoregressiven Prozesses

§ II.3 Starke Konsistenz des Kleinste-Quadrate-Schätzers

§ II.4 Asymptotische Normalität des Kleinste-Quadrate-Schätzers

§ II.5 Spezialisierung des dynamischen Modells auf den Fall unabhängiger Störvariabler

Kapitel III: Girsanovs Theorem im Hilbertraum und eine Anwendung auf ein statistisches Schätzproblem, welches mit einer Hilbertraum-wertigen stochastischen Differentialgleichung zusammenhängt

§ III.1 Girsanov-Sätze für Hilbertraum-wertige Wienerprozesse und stochastische Differentialgleichungen

§ III.2 Parameterschätzung für den Driftterm einer unendlich-dimensionalen stochastischen Differentialgleichung

Kapitel I: Einleitung

Viele in der Realität ablaufende Vorgänge beispielsweise im Bereich der natürlichen Umwelt, der Technik oder Ökonomie können quantitativ durch ein stochastisches Modell annähernd beschrieben werden. Darunter fallen u. a. Phänomene, deren Entwicklung und Veränderung im Lauf der Zeit wesentliche Merkmale sind. Die Einführung von zufälligen Modellkomponenten wird dabei meist durch die Komplexität der zu beschreibenden Vorgänge nahegelegt. Im allgemeinen kann man nicht alle auf ein System wirkenden Einflüsse exakt erfassen und modelliert deshalb die verschiedenen Einflußfaktoren mittels stochastischer Größen. Modellansätze bzw. zugehörige Datenreihen dieser Art nennt man Zeitreihen. Mathematisch handelt es sich um stochastische Prozesse bzw. um deren Realisierungen. Das Studium von Zeitreihen dient verschiedenen (praktischen) Zwecken. Ein wichtiges Gebiet ist die Vorhersage zukünftiger Prozeßverläufe aufgrund von in der Vergangenheit gemessenen Werten. Zum anderen kann auch die Steuerung eines dynamischen Systems wünschenswert sein. Um allerdings Vorhersagen aufstellen oder ein System kontrollieren zu können, bedarf es einer genauen mathematischen Beschreibung des Systems.

Die am weitesten verbreiteten Modelle der Zeitreihenanalyse sind stochastische Versionen von Differenzen- bzw. Differentialgleichungen. Für zeitdiskrete Systeme betrachtet man sogenannte autoregressive (AR(p)-) Prozesse p-ter Ordnung, die die Gestalt $x_{k+1} = f_k(\underline{\theta}; x_k, x_{k-1}, \ldots, x_{k-p+1}) + v_{k+1}$ besitzen, wobei $(x_k)_{k \in \mathbb{N}}$ den Systemprozeß bezeichnet (d. h. den diskreten stochastischen Prozeß, der das System beschreibt) und v_k die Systemstörungen, $\theta \in \Theta$ einen noch frei verfügbaren bzw. unbekannten Parameter, und $f_k : \Theta \times H^p \longrightarrow H$ sind Abbildungen mit dem Wertebereich H, dem Zustandsraum des Prozesses $(x_k)_{k \in \mathbb{N}}$. Die allgemeine strukturelle und statistische Theorie derartiger zeitdiskreter AR-Prozesse

ist u. a. in den Lehrbüchern [AN, '71], [AO, '67], [BO/JE, '70], [HA, '70] und [KA/RA, '76] dargelegt.

Für zeitstetige Systeme ist die Beschreibung mittels einer stochastischen Differentialgleichung üblich: $dx_t = f(t,\underline{\theta},x_t)\,dt + g(t,\underline{\theta},x_t)\,dw_t$. Solche Gleichungen werden u. a. in den Monographien [AR, '73], [GI/SK, '79], [FR, '75], [IK/WA, '81] untersucht. Auf statistische Aspekte dieser Prozesse gehen [LI/SH, '78] und [BA/RA, '80] ein. Man kann die obigen Modelle insofern noch flexibler gestalten, indem man annimmt, daß der Prozeß $(x_k)_{k\in\mathbb{N}}$ nicht direkt beobachtet werden kann, sondern nur mittels einer (additiv) verrauschten Meßreihe $(y_k)_{k\in\mathbb{N}}$, also $y_k = g_k(x_k) + e_k$, wobei e_k den jeweiligen Meßfehler angibt. Entsprechende Darstellungen kann man für die zeitstetige Situation anführen. Solche Gleichungen werden als Filtersysteme bezeichnet. Sie werden ausführlich in den weiter unten zitierten Büchern behandelt.

Wir stellen nun kurz einige statistische Fragestellungen bzgl. der oben erwähnten Modelle vor. Zunächst stellt sich die Frage der Modellauswahl, d. h. der Wahl einer bestimmten Modellklasse, um ein reales Phänomen bzw. einen von diesem erzeugten Datensatz adäquat zu beschreiben. Dabei geht es zum Beispiel um die Frage, ob eine Modellierung durch einen AR(p)-Prozeß mit passender Zustandsraumdimension geeignet ist. Ferner ist die Ordnung p zu bestimmen. Probleme dieser Art sind z. B. in den Kapiteln VIII und IX des Buches [KA/RA, '76] behandelt. Im Abschnitt 2.9 des Übersichtsartikels von Cox [CO, '81] findet man weitere Referenzen. Ist die Modellspezifizierung bereits vollzogen, so tauchen meist noch weitere (unbekannte) Parameter $\underline{\theta}$ in der dynamischen Gleichung auf (vgl. oben). Es kann dann wichtig sein, den wahren Parameter $\underline{\theta}$ aufgrund von gemessenen Daten annähernd zu bestimmen, d. h. zu schätzen (z. B. zwecks Verifizierung einer Theorie oder um gewisse Naturkonstanten zu ermitteln). Diese Arbeitsgebiete werden in den Ingenieurwissenschaften und in der Systemtheorie Modell-Identifikation genannt. Ein weiteres

für die Anwendung wichtiges Gebiet der Systemtheorie ist die Theorie der Filterung stochastisch gestörter Signale. Dabei geht man davon aus, daß der Systemprozeß $(x_k)_{k \in \mathbb{N}}$ selbst nicht beobachtet werden kann, sondern daß man nur über die verrauschten Meßwerte $(y_k)_{k \in \mathbb{N}}$ verfügt. Die Problematik besteht nun darin, den Zustand x_k möglichst genau zu schätzen aufgrund der Kenntnis der Vergangenheit des Beobachtungsprozesses, d. h. $y_1,\ldots,y_k$. Theoretisch ist die beste Schätzung bzgl. der Abweichung im quadratischen Mittel gegeben durch den bedingten Erwartungswert $E(x_k|y_1,\ldots,y_k)$. Die Filter-Theorie besteht nun im wesentlichen daraus, (sequentielle) Algorithmen zur Berechnung dieses Erwartungswertes anzugeben (vgl. u. a. [JA, '70], [LI/SH, '78], [AS, '79], [BA/RA, '80], [HA/WI, '81]). Werden die Aufgaben der Identifikation und der Filterung gleichzeitig bewältigt, spricht man vom adaptiven Filtern.

Die vorliegende Arbeit behandelt ausschließlich Fragen der Modell-Identifikation, und zwar Parameterschätzprobleme. Als Zustandsraum wird dabei ein separabler Hilbertraum angenommen, so daß insbesondere die Schätztheorie für Prozesse mit endlich-dimensionalem euklidischem Zustandsraum enthalten ist. Darüber hinaus bietet die Verfügbarkeit über einen unendlich-dimensionalen Zustandsraum eine Flexibilität, die sich z. B. als nützlich für die Modellierung von räumlich ausgedehnten Systemen erweisen kann, d. h. von Systemen, deren Datensätze zu einem festen Zeitpunkt aus Funktionsgraphen über einem bestimmten Gebiet bestehen.

Die am häufigsten verwendeten Verfahren zur Parameterschätzung sind die Methode der kleinsten Quadrate (engl. "least-square-method") und die Maximum-Likelihood-Methode. Im weiteren Verlauf der Arbeit werden wir beide benutzen, um geeignete Schätzformeln zu konstruieren. Deshalb stellen wir beide Verfahren hier kurz anhand einer allgemeinen statistischen Situation dar.

Die Methode der kleinsten Quadrate (die in ihrer einfachsten Form auf Gauß zurückgeht) besteht im wesentlichen darin, daß man geeignete Funktionale der Beobachtungswerte y_k (bzw. x_k) und des unbekannten Parameters $\underline{\theta}$ bildet derart, daß diese gewissen zentrierten (Fehler-) Variablen gleichen. Die aufsummierten Quadrate der Funktionale (≡ Summe der Fehlerquadrate) minimiert man dann bezüglich der Größe $\underline{\theta}$. Nach Auflösung der Extremwertgleichung erhält man dann eine Schätzformel $\hat{\underline{\theta}}_n$ für $\underline{\theta}$ als Funktional von $y_1,\ldots,y_n$. (Bezüglich der statistischen Begründung und Bedeutung dieser Methode vgl. man z. B. [LI, '61], [SC, '66] und [WI/NÖ, '70]).

Die Maximum-Likelihood-Methode (die auf R.A. Fisher zurückgeht) setzt die Existenz von Wahrscheinlichkeitsdichten des Datensatzes $(y_1,\ldots,y_n)$ (bzw. $(y_s)_{0\leq s\leq t}$ im zeitstetigen Fall) bezüglich eines festen dominierenden Maßes μ auf H^n voraus. Als Schätzwert $\hat{\underline{\theta}}_n$ für $\underline{\theta}$ wählt man einen Punkt, für den die Likelihood-Funktion $\underline{\theta} \longrightarrow L(\eta_1,\ldots,\eta_n;\underline{\theta})$ maximal wird. Der Term $L(\eta_1,\ldots,\eta_n;\underline{\theta})$ ist gerade die gemeinsame μ-Dichte von $y_1,\ldots,y_n$ bei zugrundeliegendem Parameter $\underline{\theta}$, ausgewertet für die Realisierungen $\eta_1,\ldots,\eta_n$. Dieses Vorgehen bedeutet z. B. im Fall einer diskreten Wahrscheinlichkeitsverteilung $P_{y_1,\ldots,y_n;\underline{\theta}}$ eine solche Wahl des Parameters $\underline{\theta}$, daß dem eingetretenen Ereignis $\{y_1 = \eta_1;\ldots,y_n = \eta_n\}$ größte Wahrscheinlichkeit zukommt. (Diese Methode ist u. a. in [SC, '66] und [WI/NÖ, '70] ausführlich dargestellt.)

Die Bedeutung beider Methoden besteht darin, daß sie unter bestimmten Bedingungen wünschenswerte asymptotische Eigenschaften der Schätzerfolge $(\hat{\underline{\theta}}_n)_{n\in\mathbb{N}}$ gewährleisten. Unter diesen Eigenschaften wollen wir die Konsistenz und die asymptotische Normalität hervorheben. Dabei heißt

$(\hat{\underline{\theta}}_n)_{n\in\mathbb{N}}$ (stark) konsistent, falls $\hat{\underline{\theta}}_n$ P-stochastisch (P-fast sicher) gegen den wahren Parameter $\underline{\theta}$ konvergiert mit n gegen ∞. $\hat{\underline{\theta}}_n, n\in\mathbb{N}$, ist asymptotisch normalverteilt, falls $\sqrt{n}\,(\hat{\underline{\theta}}_n - \underline{\theta})$ mit n gegen ∞ schwach gegen eine Normalverteilung konvergiert.

Im folgenden gehen wir auf die Untersuchungen in der vorliegenden Arbeit ein und diskutieren die Zeitreihen-Literatur, die sich näher mit der hier behandelten Thematik befaßt. Im zweiten Kapitel lösen wir das Problem, den Parameter $\underline{\theta} = (\theta_1,\ldots,\theta_m) \in \Theta \subset \mathbb{R}^m$ des folgenden Filter-Systems zu schätzen:

$$x_{k+1} = (A_o + \sum_{\ell=1}^{m} \theta_\ell A_\ell)\, x_k + v_{k+1},\quad y_k = Bx_k + e_k.$$

Dabei ist der Zustandsraum des Prozesses $(x_k)_{k\in\mathbb{N}}$ (des "Signalprozesses" oder "Systemprozesses") ein separabler Hilbertraum H, und die Beobachtungen y_k sind Elemente eines weiteren separablen Hilbertraums K. Die linearen Operatoren $A_\ell, \ell = 0,\ldots,m$, B sind bekannt; v_k, e_k sind stochastische Störungen, die Systemstörungen bzw. Meßfehler genannt werden. Im Abschnitt II.1 wird das System detaillierter vorgestellt. Bei diesem Modell handelt es sich also um einen linearen AR(1)-Prozeß, der nur indirekt beobachtet werden kann. Die direkte Beobachtungsmöglichkeit ist natürlich mittels der Wahl $e_k = 0$, $B = id_H$ in diesem Modell enthalten.

Letztere Situation wurde zuerst von Mann und Wald untersucht, [MA/WA, '43]. Die beiden Autoren behandeln stationäre, r-dimensionale AR(p)-Prozesse und leiten die Schätzformel mittels der Maximum-Likelihood-Methode her, indem sie zunächst annehmen, daß die Systemstörungen normalverteilt sind. Anschließend übernehmen sie die gleiche Schätzformel auch für nicht-gaußsche, unabhängige, identisch verteilte Störvariable. Unter der Annahme der Existenz aller Momente der Störungen weisen sie die (schwache) Konsistenz und

asymptotische Normalität des Schätzers nach. Zeitgemäße Darstellungen der asymptotischen Theorie der Schätzer für dieses Problem findet man u. a. in den Lehrbüchern von T.W. Anderson und Basawa/Rao, [AN, '71], [BA/RA, '80].

In neueren Arbeiten wird versucht, die Voraussetzung der Unabhängigkeit der Störvariablen abzuschwächen. So leiten Grosspietsch und Graupe [GR/GR, '81] für einen skalaren (direkt beobachtbaren) AR(p)-Prozeß die starke Konsistenz des Least-Square-Schätzers her u. a. für den Fall, daß die Störungen eine stationäre Martingaldifferenzfolge bilden. Bustos [BU, '82] (Näheres s. u.) arbeitet mit φ-mischenden Störungen.

Untersuchungen über die Parameterschätzung bei gestörten Messungen (d. h. bei indirekter Beobachtung) sind weniger zahlreich und neueren Ursprungs. W.N. Anderson et al., [AN/KL/WO, '69], betrachten ein stationäres, endlichdimensionales System $x_{k+1} = A\,x_k + v_k$, $y_k = x_k + e_k$ mit unabhängigen Störungen v_k, e_k. Sie konstruieren ausgehend von der Beziehung $E(y_k y_{k-2}^*) = AE(y_{k-1}y_{k-2}^*)$ eine Schätzerfolge $\hat{A}_n$ für A, $n \in \mathbb{N}$, und weisen deren starke Konsistenz nach.

Pagano [PA, '74] studiert einen skalaren AR(p)-Systemprozeß. Sowohl die Systemstörungen als auch die Meßfehler sind gaußsche Variable. Paganos Schätzmethode besteht darin, daß er für die Beobachtungen eine Darstellung als ARMA(p,p)-Prozeß findet (ARMA = autoregressive moving average) mit den gleichen AR-Koeffizienten wie bei dem ursprünglichen Signalprozeß. Dabei sind ARMA(p,q)-Prozesse solche AR(p)-Systeme, bei denen die Störungen v_k aus gleitenden Durchschnitten (der Länge q) von unabhängigen, identisch verteilten Variablen bestehen. Das obige Schätzproblem ist also reduziert auf die Identifikation eines ARMA-Prozesses, der direkt beobachtet werden kann

und somit mit bekannten Methoden lösbar. Ähnlich wie Pagano gehen Kashyap und Rao ([KA/RA, '76], S. 75 ff.) vor. Die Autoren untersuchen skalare ARMA-Prozesse samt verrauschter Messungen und leiten für den Beobachtungsprozeß wiederum eine ARMA-Darstellung her, in der die ursprünglichen Koeffizienten wieder auftauchen.

Recht allgemeine r-dimensionale Modelle der Form $z(n) = \sum_{j=0}^{\infty} C(j;\underline{\theta})\ \varepsilon\ (n-j)$ mit unkorrelierten $\varepsilon(\ell)$ betrachtet Dunsmuir [DU, '79].
Die Situation der indirekten Beobachtung erweist sich als mit dieser Darstellung verträglich. Der Autor konstruiert seinen Schätzer mittels der gaußschen Likelihood-Funktion, obwohl die Systemvariablen nicht normalverteilt zu sein brauchen. Dunsmuir behandelt seine Zeitreihen mit der diskreten Fourier-Transformation und arbeitet anschließend im Frequenzbereich. Es wird die asymptotische Normalität der Schätzer nachgewiesen, obwohl diese nicht explizit gegeben sind, sondern nur als Extremwerte entsprechender Likelihood-Funktionale festgelegt sind.

Ebenfalls 1979 veröffentlichte Uosaki zwei Arbeiten über sequentielle Parameterschätzung bei Filter-Systemen, [UO, '79;1], [UO, '79;2].
Er untersucht eindimensionale AR(p)-Prozesse mit gestörten Beobachtungen. Der Autor gibt mittels der Methode der stochastischen Approximation eine sequentielle Schätzerfolge an, deren starke Konsistenz er verifiziert. Die stochastische Approximation ist dabei eine wahrscheinlichkeitstheoretische Version des Newton-Raphson-Verfahrens in der Numerik (s. z. B. [WA, '69]).

Einige Arbeiten neueren Datums beschäftigen sich mit der Parameterschätzung bei skalaren autoregressiven Prozessen unter dem Gesichtspunkt der Robustifizierung der Schätzer. Dabei werden die Beobachtungen y_k als statistische "Ausreißer" (engl. "outliers") bzgl. der Systemgrößen x_k interpretier

So studieren Denby und Martin [DE/MA, '79] das einfache eindimensionale System $x_{k+1} = \theta x_k + v_{k+1}$, $y_k = x_k + e_k$. Sie gelangen zu einer Schätzformel für θ, indem sie Terme der Gestalt $\sum_{k=1}^{n-1} L(y_{k+1} - \theta y_k)$ bzgl. θ minimieren, wobei L eine näher zu bestimmende symmetrische, robustifizierende Verlustfunktion ist. Im allgemeinen muß $L(z) \neq z^2$ sein, da ansonsten die Schätzformel θ_n als Funktional von $y_1,\ldots,y_n$ identisch mit dem Funktional der $x_1,\ldots,x_n$ für den Fall der direkten Beobachtungsmöglichkeit wäre. Also würde i. allg. natürlich keine Konsistenz des Schätzers mehr vorliegen (zu letzterem Aspekt vgl. man auch die Arbeit von Griffith [GRI, '82]). Die zur Zeit umfassendste Untersuchung auch der asymptotischen Eigenschaften dieser so gewonnenen Schätzer stammt von Bustos, [BU, '82]. In dieser Arbeit werden skalare AR(p)-Prozesse $(x_k)_{k\in\mathbb{N}}$ betrachtet zusammen mit einer sogenannten Kontaminationsgleichung $y_k = (1 - u_k)x_k + u_k e_k$, wobei der Prozeß u_k den Anteil der Störung e_k festlegt. Sämtliche Störprozesse sind einzeln φ-mischend, jedoch unabhängig voneinander.

Als nächsten Punkt wollen wir die Frage nach einer gewissen Optimalität unserer bzw. der in der Literatur verwendeten Schätzer diskutieren. Dabei beschränken wir uns auf den Begriff der asymptotischen Effizienz. Wir erläutern zunächst diesen Begriff in der Situation der klassischen Statistik, d. h. bei Vorliegen von Realisierungen unabhängiger, identisch verteilter Variabler x_k, $k\in\mathbb{N}$. Zusätzlich muß noch die Existenz von Dichten $f(x,\underline{\theta}),\underline{\theta}\in\Theta$, der zugehörigen Verteilungen bzgl. eines gemeinsamen dominierenden Maßes gefordert werden, die außerdem gewisse Regularitätseigenschaften besitzen müssen. Man nennt dann eine Schätzerfolge $(\hat{\underline{\theta}}_n)_{n\in\mathbb{N}}$ asymptotisch effizient, wenn sie asymptotisch normalverteilt ist und die Kovarianz der Grenzverteilung, $S(\underline{\theta})$, mit der Inversen der Fisherschen Informationsmatrix $I(\underline{\theta})$ übereinstimmt für alle $\theta\in\Theta$.

Dabei ist $I(\underline{\theta})$ definiert als $(E(\frac{\partial \ln f(x_1,\underline{\theta})}{\partial\theta_i} \cdot \frac{\partial \ln f(x_1,\underline{\theta})}{\partial\theta_j}))_{i,j=1,\ldots,m}$.

Diese Definition der asymptotischen Effizienz ist motiviert durch ein Resultat von Bahadur, [BAH, '67] (vgl. auch [WI/NÖ, '70], S. 72), welches unter gewissen Zusatzvoraussetzungen die Ungleichung $S(\underline{\theta}) \geq (I(\underline{\theta}))^{-1}$ sicherstellt.

Beim Studium der Zeitreihen liegt eine wesentlich komplexere Situation vor. So zeigen Akahira und Takeuchi im Abschnitt 3.2 ihrer Monographie [AK/TA, '81], daß der Least-Square-Schätzer bzgl. eines einparametrigen, eindimensionalen AR(1)-Prozesses mit direkter Beobachtungsmöglichkeit dann und nur dann asymptotisch effizient ist, wenn die (unabhängigen) Systemstörungen normalverteilt sind. In der allgemeinen Situation, die in dieser Arbeit behandelt wird, kann man also keine asymptotische Effizienz erwarten. Selbst eine untere Schranke für die Kovarianz der gaußschen Grenzverteilung als Funktion der Informationsmatrix ist nicht erhältlich, da letztere in unserem Fall gar nicht konstruiert werden kann. Im allgemeinen existieren nämlich keine Dichten bzgl. eines gemeinsamen dominierenden Maßes für Verteilungen im Hilbertraum, da z. B. schon gaußsche Maße in der Regel senkrecht aufeinander stehen (vgl. z. B. [KU, '75], § II.3). Effizienzfragen bei endlich-dimensionalen Systemen haben auch Denby und Martin, [DE/MA, '79], [MA, '79], [MA, '82] und Pagano, [PA, '74] diskutiert.

Wir skizzieren nun den Inhalt unseres zweiten Kapitels. Nachdem im ersten Abschnitt die mathematischen Details des Filtersystems festgelegt werden, konstruieren wir im Abschnitt II.2 den Parameterschätzer $\hat{\underline{\theta}}_n$ mittels einer modifizierten Methode der kleinsten Quadrate. In den weiteren Abschnitten diskutieren wir das asymptotische Verhalten dieses Schätzers. Wir beweisen zunächst die starke Konsistenz von $\hat{\underline{\theta}}_n$ (s. Satz II.3.1), wobei im wesentlichen starke Gesetze der großen Zahlen für ergodische Prozesse als Hilfsmittel

herangezogen werden. Anschließend (s. Satz II.4.1) leiten wir die asymptotische Normalität des Schätzers her, indem wir einen Zentralen Grenzwertsatz für Funktionale von φ-mischenden Folgen benutzen. Im Abschnitt II.5 behandeln wir den Spezialfall unabhängiger Störungen und geben u. a. explizite Formeln für die Kovarianz der Grenzverteilung (der Folge $\sqrt{n}(\hat{\theta}_n - \theta)$) an (vgl. Satz II.5.2).

Abschließend kommentieren wir das zeitstetige Identifikationsproblem des dritten Kapitels, in dem wir mit einem Hilbertraum-wertigen Prozeß arbeiten, der eine stochastische Differentialgleichung löst. Parameterschätzungen für die Driftterme endlich-dimensionaler stochastischer Differentialgleichungen sind u. a. ausführlich dargestellt in den Monographien von Liptser/Shiryayev [LI/SH, '78], S. 202 ff., und Basawa/Rao [BA/RA, '80], S. 210 ff. (s. auch die dort zitierte Literatur und u. a. die Originalarbeiten von Le Breton und Lanska, [BR, '74], [LA, '79]). Endlich-dimensionale Systeme im Fall indirekter Beobachtung sind von Bagchi und Borkar [BA, '78], Tugnait [TU, '80] und Makhno [MAK, '80] untersucht worden. Ebenfalls von Bagchi und Borkar [BA/BO, '81] stammt eine Arbeit über ein Identifikationsproblem bzgl. eines unendlich-dimensionalen Systems. Die zugehörigen Beobachtungen sind aber endlich-dimensional, so daß man die unbekannten Parameter mit Hilfe der endlich-dimensionalen Maximum-Likelihood-Theorie schätzen kann.

Wir dagegen arbeiten (im Kap. III) direkt mit Prozessen $(z(t))_{t \in [0,\infty)}$, die durch Hilbertraum-wertige stochastische Differentialgleichungen $dz(t) = \theta Az(t)dt + dw(t)$ gegeben sind. Dabei ist A infinitesimale Erzeugende einer stark stetigen Operatorenhalbgruppe auf dem reellen, separablen Hilbertraum H und $(w(t))_{t \in [0,\infty)}$ ein H-wertiger Wienerprozeß (vgl. Def. III.1.1).

Wir benutzen diejenige Theorie Hilbertraum-wertiger stochastischer Differentialgleichungen, die (z. B.) in dem Lehrbuch von Curtain und Pritchard [CU/PR, '78] dargestellt ist.

Für die Konstruktion unseres Maximum-Likelihood-Schätzers benötigen wir unendlich-dimensionale Versionen der Girsanov-Theoreme. Sätze vom Girsanov-Typ befassen sich zum einen mit der Transformation von Wienerprozessen bei gleichzeitiger geeigneter Veränderung des zugrundeliegenden Wahrscheinlichkeitsmaßes und zum andern mit daraus abgeleiteten Ergebnissen über die Absolutstetigkeit von Maßen, die durch Lösungen stochastischer Differentialgleichungen mit verschiedenen Drifttermen induziert werden. Die Situation des endlich-dimensionalen Zustandsraums ist u. a. ausführlich in den Büchern von Friedman und Liptser, Shiryayev, [FR, '75], [LI/SH, '77] dargestellt. Resultate dieser Art für Hilbertraum-wertige Prozesse sind schon von Ouvrard [OU, '72], [OU, '73] erarbeitet worden. Er verwendet dabei das Ito-Lemma und abstrakte Martingaltheorie jeweils für Hilbertraum-wertige stochastische Prozesse. Unsere Herleitung im Abschnitt III.1 ist direkter, indem wir eine Projektionsmethode verwenden, die erlaubt, das Problem auf den Fall endlich-dimensionaler Zustandsräume zurückzuführen. Diese Vorgehensweise ist ähnlich der Methode Bells [BE], der allerdings das entsprechende Problem im Rahmen des abstrakten Wienerraums im Sinne von Gross (vgl. [GR, '65]) behandelt.

Schon 1967 diskutierte Daletskii unendlich-dimensionale stochastische Differentialgleichungen (allerdings in einem anderen formalen Rahmen) und formulierte einen Satz über die Absolutstetigkeit von Maßen bzgl. verschiedener Lösungen (s. [DA, '67], Th. 2.5). Ein abstraktes Girsanov-Theorem für H-wertige Prozesse findet man auch im Buch von Metivier und Pellaumail [ME/PE, '80], S. 56. Ein weiteres Resultat in dieser Richtung erzielten Hitsuda und H. Watanabe [HI/WA, '76], § 4.

Im zweiten Abschnitt des Kapitels III konstruieren wir schließlich mittels der im Abschnitt III.1 erzielten Ergebnisse einen Maximum-Likelihood-Schätzer für den in der Differentialgleichung auftretenden Parameter θ und weisen dessen starke Konsistenz und (unter zusätzlichen Voraussetzungen) asymptotische Normalität nach (s. Sätze III.2.1 und III.2.2).

Ich möchte Herrn Prof. Dr. S. Albeverio für viele hilfreiche Diskussionen sowie für die Betreuung der vorliegenden Arbeit danken. Frau Richter und Frau Mischke danke ich für die sorgfältige Reinschrift des Manuskripts.

Kapitel II: Parameterschätzung für autoregressive Prozesse bei indirekter Beobachtung

§ II.1: Erläuterung des Modells

Im Kapitel II der vorliegenden Arbeit diskutieren wir statistische Schätzprobleme, die mit folgender Prozeß-Dynamik zusammenhängen:

$$(I) \quad x_{k+1} = (A_0 + \sum_{\ell=1}^{m} \theta_\ell A_\ell) x_k + v_{k+1},$$

$$(II) \quad y_k = Bx_k + e_k.$$

Dabei ist $k \in Z_+$ der diskrete Zeitparameter. Gleichung (I) beschreibt die zeitliche Entwicklung des Systemprozesses $(x_k)_{k \in Z_+}$ und (II) legt die verrauschten Beobachtungen y_k, $k \in Z_+$, fest. Die Signale x_k nehmen Werte in einem separablen Hilbertraum H mit Skalarprodukt $\langle \cdot, \cdot \rangle_H$ an. x_0 ist eine beliebige H-wertige Zufallsgröße endlicher Varianz. Die zufällige Entwicklung des Signalprozesses $(x_k)_{k \in Z_+}$ wird erzeugt durch die Folge $(v_k)_{k \in \mathbb{N}}$, einer Folge von H-wertigen Zufallsvariablen, den Systemstörungen, welche auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ definiert sind. Die implizit vorausgesetzte Meßbarkeit der v_k bezieht sich natürlich auf die Borel-σ-Algebra in H. Auf die stochastische Struktur der v_k, $k \in \mathbb{N}$, gehen wir weiter unten näher ein.

Die Koeffizienten A_ℓ, $\ell = 0,1,\ldots,m$ sind Elemente des Banachraumes $(\mathcal{L}(H), \|\cdot\|_{\mathcal{L}(H)})$ aller beschränkten linearen Operatoren auf H. Sie werden als bekannt vorausgesetzt im Gegensatz zu den reellen Parametern θ_ℓ ("Systemparameter"), $\ell = 1,\ldots,m$, die mittels der Meßwerte y_k, $k \in Z_+$, geschätzt werden sollen. Damit das System (I) eine stationäre Lösung besitzt, stellen wir folgende Stabilitätsbedingung auf:

(S) $\|A(\underline{\theta})\|_{\mathcal{L}(H)} < 1$ für den (wahren) Parameter $\underline{\theta} = (\theta_1,\dots,\theta_m) \in \mathbb{R}^m$,
wobei $A(\underline{\theta}) := A_o + \sum_{\ell=1}^{m} \theta_\ell A_\ell$.

Der Beobachtungsprozeß $(y_k)_{k\in Z_+}$ besitzt Werte in einem separablen Hilbertraum $(K, \langle\cdot,\cdot\rangle_K)$. B ist ein bekannter beschränkter linearer Operator, der H in K abbildet. Die Beobachtungsfehler $(e_k)_{k\in Z_+}$ sind K-wertige Zufallsgrößen auf $(\Omega,\mathcal{A},P)$.

Wir legen nun die stochastische Struktur der Störungsprozesse $(v_k)_{k\in\mathbb{N}}$ und $(e_k)_{k\in Z_+}$ fest, wobei wir (o.E.) annehmen, daß die Störungen auch für negative Zeiten k erklärt sind. Letzteres erleichtert uns später die Konstruktion stationärer Varianten des Signal- und des Beobachtungsprozesses.

Zuerst verlangen wir die strikte Stationarität des Prozesses $(v_k,e_k)_{k\in Z}$,[1] welcher auf $(\Omega,\mathcal{A},P)$ definiert ist und Werte in $H\times K$ annimmt. Weiterhin sei der Prozeß $(v_k,e_k)_{k\in Z}$ φ-mischend im Sinne der Monographie von Billingsley [BI, '68], S. 166, (vgl. auch [IB/LI, '71], dort lautet die Bezeichnung "uniform mixing", siehe Def. 17.2.2). Bevor wir die Eigenschaft eines Prozesses, φ-mischend zu sein, inhaltlich erläutern, geben wir eine präzise Definition. Dazu sei $(\eta_k)_{k\in Z}$ ein Prozeß auf $(\Omega,\mathcal{A},P)$ mit Werten in einem beliebigen meßbaren Raum $(X,\mathcal{B})$. Für $a,b\in\overline{Z}$, $a\le b$, setze man $\mathcal{M}_a^b := \mathcal{M}_a^b(\eta_k)_{k\in Z} := \sigma(\eta_k : k\in Z : a\le k\le b)$, d.h. $\mathcal{M}_a^b$ ist die σ-Algebra, die von den Zufallselementen $\eta_a,\dots,\eta_b$ erzeugt wird. Man definiert nun:

$(\eta_k)_{k\in Z}$ ist φ-mischend genau dann, wenn

$\varphi_n := \sup\{|P(E_2|E_1) - P(E_2)| : E_1\in\mathcal{M}_{-\infty}^{k},\ E_2\in\mathcal{M}_{k+n}^{\infty}, k\in Z\}$

gegen 0 konvergiert mit $n\to\infty$. ($P(E_2|E_1) := 0$, falls $P(E_1) = 0$).—

Der Begriff "φ-mischend" stellt eine Erweiterung des Begriffes der

---

[1] Die strikte Stationarität eines stochastischen Prozesses (bzw. Stationarität im engeren Sinne) ist z.B. in [MÜ, '75], S. 228, definiert.

stochastischen Unabhängigkeit dar, denn für eine Folge von unabhängigen Zufallsvariablen ist $\varphi_n = 0$ für alle $n \in \mathbb{N}$. Natürlich umfaßt dieser Begriff auch q-abhängige Variablenfolgen, insbesondere sind deshalb endliche gleitende Durchschnitte von unabhängigen Variablen φ-mischend. Die Mischungseigenschaft bedeutet, daß mit wachsender Zeitspanne n die Prozeßvariablen η_k und η_{k+n} sich beliebig wenig stochastisch beeinflussen, d.h. asymptotisch unabhängig sind. Ihre Bedeutung verdankt die Mischungseigenschaft vor allem der Tatsache, daß für φ-mischende Variablenfolgen (und auch für von ihnen gebildete Funktionale) noch zentrale Grenzwertsätze Gültigkeit haben, also asymptotische Normalität für die standardisierten Partialsummen gesichert ist (vgl. [Bl, '68], § 20,21).

Wir bemerken noch (ohne Beweis), daß die strikte Stationarität und die Mischungseigenschaft des zweikomponentigen Prozesses $(v_k, e_k)_{k \in Z}$ auf jeden Fall dann gegeben ist, wenn die Prozesse $(v_k)_{k \in Z}$ und $(e_k)_{k \in Z}$ einzeln strikt stationär und φ-mischend und voneinander unabhängig sind.

Wir benötigen folgende Momentenbedingungen für die Störungen: v_k und e_k sind zentriert und die zweiten Momente existieren; d.h. $E(\|v_o\|_H^2) = \operatorname{tr} S_v < \infty$ und $E(\|e_o\|_K^2) = \operatorname{tr} S_e < \infty$, wobei S_v bzw. S_e die den Verteilungen $P \bullet v_o$ bzw. $P \bullet e_o$ zugeordneten Kovarianzoperatoren darstellen; $\operatorname{tr}(\cdot)$ bezeichnet die Spur eines (nuklearen) Operators (siehe [KU, '75], Th. 2.1). Außerdem besitze die Startgröße x_o ein endliches zweites Moment, $E(\|x_o\|_H^2) < \infty$.

Aus Gründen, die erst im Abschnitt II.2 (Konstruktion des Schätzers) ersichtlich werden, muß noch gefordert werden, daß folgende Erwartungswerte nicht von dem unbekannten Parameter $\underline{\theta}$ abhängen:

$E(\langle C_\ell \tilde{y}_o, C_j e_o \rangle)$ und $E(\langle C_\ell \tilde{y}_o, Bv_l + e_l - C_o e_o \rangle)$.

Hierbei ist $\tilde{y}_o$ eine stationäre Variante des Meßwertes y_o, die zu Beginn des Abschnitts II.3 mathematisch definiert wird. Wie man sich leicht überlegen kann, hängen die oben aufgeführten Erwartungswerte insbesondere dann nicht von $\underline{\theta}$ ab, wenn $\tilde{x}_o$ (die stationäre Variante von x_o, vgl. Abschnitt II.3) oder auch $(v_\ell)_{-\infty<\ell\leq -1}$ stochastisch unabhängig von (v_1, e_o, e_1) sind. Diese Situation liegt demnach auf jeden Fall dann vor, wenn $(v_k)_{k\in Z}$ und $(e_k)_{k\in Z}$ jeweils eine Folge von unabhängigen, identisch verteilten Zufallsgrößen bilden, die zusätzlich als Folgen voneinander unabhängig sind. Letzteres Modell wird gesondert im Abschnitt II.5 behandelt.

Abschließend müssen wir noch eine Beobachtbarkeitsbedingung an das System (I),(II) stellen:

(O) Für alle $\ell = 0,1,\ldots,m$ existieren Operatoren $C_\ell \in \mathcal{L}(K)$ derart, daß $C_\ell B = BA_\ell$ ist.

Eine einfache Rechnung zeigt, daß (O) äquivalent zur folgenden Bedingung ist:

(O') Für alle $\ell = 0,1,\ldots,m$ existieren $C_\ell \in \mathcal{L}(K)$ derart, daß $C(\underline{\theta})B = BA(\underline{\theta})$ für alle $\underline{\theta}\in\mathbb{R}^m$ ist, wobei $C(\underline{\theta}) := C_o + \sum_{\ell=1}^{m} \theta_\ell C_\ell$.

Eine notwendige Voraussetzung für die Bedingung (O) ist folgende Relation:

(+) Kern $(B) \subset$ Kern (BA_ℓ) für alle $\ell = 0,1,\ldots,m$.

Falls B invertierbar und $B^{-1} \in \mathcal{L}(K,H)$ ist, existieren die Operatoren C_ℓ, $\ell = 0,1,\ldots,m$ und $C_\ell = BA_\ell B^{-1}$. Im Falle einer endlich-dimensionalen Beobachtung, d.h. dim $K < \infty$, ist die Bedingung (+) auch hinreichend für (O), da man C_ℓ durch folgende Festsetzung eindeutig definieren kann: $C_\ell \equiv 0$ auf dem orthogonalen Komplement des Wertebereichs von B und andererseits $C_\ell y = BA_\ell x$, falls $y = Bx$.

Übrigens ist in der einparametrigen, eindimensionalen Situation, d.h. $x_{k+1} = \theta a x_k + v_{k+1}$, $y_k = bx_k + e_k$, $\theta,a,b\in\mathbb{R}$, die Beobachtbarkeitsbedingung (O) stets erfüllt. Im folgenden schreiben wir kurz $\langle\cdot,\cdot\rangle$ für $\langle\cdot,\cdot\rangle_K$ und $\|\cdot\|$ für $\|\cdot\|_K$.

§ II.2: Herleitung eines modifizierten Kleinste-Quadrate-Schätzers für die Systemparameter des autoregressiven Prozesses

In diesem Abschnitt entwickeln wir eine Schätzformel $\hat{\underline{\theta}}_n$ für $\underline{\theta} = (\theta_1,\ldots,\theta_m)$ in die nur die Beobachtungen y_k bis zum Zeitpunkt n eingehen. Indem wir nacheinander die Gleichungen (II), (O') und (I) anwenden, erhalten wir eine Beziehung zwischen den Beobachtungsgrößen y_{k-1} und y_k, in der zusätzlich der Parameter $\underline{\theta}$ auftaucht:

$$y_k - C(\underline{\theta})(y_{k-1}-e_{k-1}) = Bv_k + e_k, \quad k\in\mathbb{N}.$$

Bv_k+e_k stellt eine zentrierte Fehlergröße dar, so daß es vernünftig erscheint, folgenden Ausdruck bzgl. $\underline{\theta}$ zu minimieren:

$$g_n(\underline{\theta}) := \sum_{k=1}^{n} \|y_k-C(\underline{\theta})(y_{k-1}-e_{k-1})\|^2$$

(Methode der kleinsten Fehler-Quadrate!). Wenn man das Gleichungssystem $\frac{\partial}{\partial\theta_\ell} g_n(\underline{\theta}) = 0$, $\ell = 1,\ldots,m$, nach $\underline{\theta}$ auflöst, erhält man folgende m-dimensionale Beziehung:

$$\tilde{\Gamma}(n)\ \hat{\underline{\theta}}_n^{(0)} = \tilde{\underline{\eta}}(n).$$

Dabei ist $\tilde{\Gamma}(n)$ eine $m\times m$-Matrix mit den Elementen

$$\tilde{\gamma}_{\ell,j}(n) = \sum_{k=1}^{n} \langle C_\ell(y_{k-1}-e_{k-1}), C_j(y_{k-1}-e_{k-1})\rangle.$$

Die Komponenten des Vektors $\tilde{\underline{\eta}}(n)$ lauten

$$\tilde{\eta}_\ell(n) = \sum_{k=1}^{n} \langle C_\ell(y_{k-1}-e_{k-1}), y_k - C_o(y_{k-1}-e_{k-1})\rangle.$$

$\hat{\underline{\theta}}_n^{(0)}$ ist ein provisorischer Kleinste-Quadrate-Schätzer (engl.: Least-Square-estimator), den wir noch abändern müssen, da in den Größen $\tilde{\gamma}_{\ell,j}(n)$ und $\tilde{\eta}_\ell(n)$ die Beobachtungsstörungen e_k auftauchen, die dem Beobachter nicht zugänglich sind.

Diese Modifikation des Schätzers besteht darin, daß man Terme, in denen die e_k auftreten, durch geeignete (von $\underline{\theta}$ unabhängige) Konstanten, hier geeignete Erwartungswerte, ersetzt in der Weise, daß der endgültige Schätzer bzw. die ihn bestimmenden Größen $\Gamma(n)$ und $\underline{\eta}(n)$ geeignetes Konvergenzverhalten insofern aufweisen, als wünschenswerte asymptotisch-statistische Eigenschaften des Schätzers gesichert sind. Für den modifizierten Least-Square-Schätzer $\hat{\underline{\theta}}_n$ erhält man dann folgende Normalengleichung:

$$\Gamma(n)\hat{\underline{\theta}}_n = \underline{\eta}(n) ,$$

wobei $\Gamma(n) = (\gamma_{\ell,j}(n))_{\substack{\ell=1,\dots,m \\ j=1,\dots,m}}$,

$$\gamma_{\ell,j}(n) = \sum_{k=1}^{n} [<C_\ell y_{k-1}, C_j y_{k-1}> - E(<C_\ell \tilde{y}_o, C_j e_o>)]$$

und $$\eta_\ell(n) = \sum_{k=1}^{n} [<C_\ell y_{k-1}, y_k - C_o y_{k-1}> - E(<C_\ell \tilde{y}_o, Bv_1 + e_1 - C_o e_o>)].$$

Dabei ist $\tilde{y}_o$ eine stationäre Variante von y_o; auf die Existenz und genaue Gestalt von $\tilde{y}_o$ gehen wir zu Beginn des Abschnitts II.3 ein. Im Abschnitt II.1 über die Modellannahmen wurde gerade gefordert, daß die obigen Erwartungswerte nicht von dem unbekannten, zu schätzenden Parameter $\underline{\theta}$ abhängen. Im Prinzip sind diese Erwartungswerte also aufgrund der sonstigen Modellparameter (wie z.B. die Kovarianzen der Störungen) bestimmbar; vgl. hierzu auch den Abschnitt II.5.

Natürlich ist $\hat{\underline{\theta}}_n = \hat{\underline{\theta}}_n(\omega)$, $\omega \in \Omega$, nur dann eindeutig durch die Normalengleichung bestimmt, wenn $\Gamma(n)(\omega)$ invertierbar ist. Später (u.a. in Satz II.3.1) geben wir hinreichende Bedingungen an, die zumindest für große $n \in \mathbb{N}$ die Invertierbarkeit von $\Gamma(n)$ P-fast sicher gewährleisten.

§ II.3: Starke Konsistenz des Kleinste-Quadrate-Schätzers

Bevor wir die starke Konsistenz des Schätzers $\hat{\underline{\theta}}_n$, d.h. $\hat{\underline{\theta}}_n \to \underline{\theta}$ P-fast sicher mit $n \to \infty$, zeigen können, benötigen wir einige vorbereitende Resultate. Zunächst führen wir dazu stationäre Varianten $\tilde{x}_k$ bzw. $\tilde{y}_k$ von x_k bzw. y_k ein. Für $q \in \mathbb{N}$ bezeichne $\tilde{x}_k^{(q)} := \sum_{\ell=0}^{q} A(\underline{\theta})^\ell v_{k-\ell}$ und $\tilde{y}_k^{(q)} := B\tilde{x}_k^{(q)} + e_k$ für alle $k \in Z$. Weiterhin betrachte man $\tilde{x}_k := \lim_{q\to\infty} \tilde{x}_k^{(q)} = \sum_{\ell=0}^{\infty} A(\underline{\theta})^\ell v_{k-\ell}$, wobei der Limes im Raum $\mathcal{L}_2(\Omega,A,P;H)$ existiert, da $\|A(\underline{\theta})\|_{\mathcal{L}(H)} < 1$ und $E(\|v_\ell\|_H^2) < \infty$ ist. Außerdem sei $\tilde{y}_k := B\tilde{x}_k + e_k, k \in Z$. Aufgrund der strikten Stationarität von $(v_k,e_k)_{k\in Z}$ sind offensichtlich auch die Prozesse $(\tilde{x}_k)_{k\in Z}$ und $(\tilde{y}_k)_{k\in Z}$ strikt stationär. Es gilt nun folgendes

Lemma II.3.1:

Für alle $\ell,j = 1,\ldots,m$ konvergiert

$$\frac{1}{n}\sum_{k=1}^{n} \langle C_\ell \tilde{y}_{k-1}, C_j \tilde{y}_{k-1}\rangle \quad \text{P-fast sicher gegen}$$

$$E(\langle C_\ell \tilde{y}_o, C_j \tilde{y}_o\rangle) \text{ mit } n \to \infty.$$

Beweis:

a) Zunächst sei bemerkt, daß $E(\langle C_\ell \tilde{y}_o, C_j \tilde{y}_o\rangle)$ existiert.

Es gilt nämlich:

$$E(|\langle C_\ell \tilde{y}_o, C_j \tilde{y}_o\rangle|) \le E(\|C_\ell \tilde{y}_o\| \cdot \|C_j \tilde{y}_o\|) \le \|C_\ell\|_{\mathcal{L}(K)} \cdot \|C_j\|_{\mathcal{L}(K)} \cdot E(\|\tilde{y}_o\|^2),$$

$$E(\|\tilde{y}_o\|^2) \le E((\|B\tilde{x}_o\| + \|e_o\|)^2) \quad \text{und}$$

$$E(\|\tilde{x}_o\|_H^2) = E(\|\sum_{\ell=0}^{\infty} A(\underline{\theta})^\ell v_{-\ell}\|^2) \le E(\|v_o\|_H^2) \cdot \sum_{\ell=0}^{\infty} \|A(\underline{\theta})\|_{\mathcal{L}(H)}^{2\ell} < \infty.$$

b)Da $\tilde{\beta}_{\ell,j}(k) := \langle C_\ell \tilde{y}_{k-1}, C_j \tilde{y}_{k-1} \rangle$, $k \in \mathbb{N}$, ein strikt stationärer Prozeß ist, wissen wir, daß das ergodische Mittel $\frac{1}{n} \sum_{k=1}^{n} \tilde{\beta}_{\ell,j}(k)$ P-fast sicher konvergiert mit $n \to \infty$ (vgl. [DO,'53], Th. 2.1, S. 465), allerdings können wir den Limes noch nicht mit $E(\tilde{\beta}_{\ell,j}(1)) = E(\langle C_\ell \tilde{y}_o, C_j \tilde{y}_o \rangle)$ identifizieren. Weil aber P-stochastische Limiten P-fast sicher bestimmt sind, reicht es zu zeigen:

(*) $\frac{1}{n} \sum_{k=1}^{n} \tilde{\beta}_{\ell,j}(k) \xrightarrow[n\to\infty]{} E(\langle C_\ell \tilde{y}_o, C_j \tilde{y}_o \rangle)$ P-stochastisch.

c) Mit der Bezeichnung $\tilde{\beta}^{(q)}_{\ell,j} := \langle C_\ell \tilde{y}^{(q)}_{k-1}, C_j \tilde{y}^{(q)}_{k-1} \rangle$, $q \in \mathbb{N}$, spalten wir wie folgt auf:

$$\frac{1}{n} \sum_{k=1}^{n} \tilde{\beta}_{\ell,j}(k) - E(\langle C_\ell \tilde{y}_o, C_j \tilde{y}_o \rangle)$$

$$= \Big[\frac{1}{n} \sum_{k=1}^{n} \tilde{\beta}_{\ell,j}(k) - \frac{1}{n} \sum_{k=1}^{n} \tilde{\beta}^{(q)}_{\ell,j}(k) \Big] + \Big[\frac{1}{n} \sum_{k=1}^{n} \tilde{\beta}^{(q)}_{\ell,j}(k) - E(\langle C_\ell \tilde{y}^{(q)}_o, C_j \tilde{y}^{(q)}_o \rangle) \Big]$$

$$+ \big[E(\langle C_\ell \tilde{y}^{(q)}_o, C_j \tilde{y}^{(q)}_o \rangle) - E(\langle C_\ell \tilde{y}_o, C_j \tilde{y}_o \rangle) \big].$$

Im folgenden schätzen wir die Terme der rechten Gleichungsseite einzeln nach oben ab.

i) Es ist $\frac{1}{n} \sum_{k=1}^{n} \tilde{\beta}_{\ell,j}(k) - \frac{1}{n} \sum_{k=1}^{n} \tilde{\beta}^{(q)}_{\ell,j}(k)$

$$= \frac{1}{n} \sum_{k=1}^{n} \big(\langle C_\ell (\tilde{y}_{k-1} - \tilde{y}^{(q)}_{k-1}), C_j \tilde{y}_{k-1} \rangle + \langle C_\ell \tilde{y}^{(q)}_{k-1}, C_j (\tilde{y}_{k-1} - \tilde{y}^{(q)}_{k-1}) \rangle \big)$$

und damit wegen der Stationarität von $\tilde{x}_k$, $\tilde{x}^{(q)}_k$, $\tilde{y}_k$ und $\tilde{y}^{(q)}_k$

$$E\Big(\Big| \frac{1}{n} \sum_{k=1}^{n} \tilde{\beta}_{\ell,j}(k) - \frac{1}{n} \sum_{k=1}^{n} \tilde{\beta}^{(q)}_{\ell,j}(k) \Big| \Big)$$

$$\leq \frac{1}{n} \sum_{k=1}^{n} \big(E(|\langle C_\ell B(\tilde{x}_{k-1} - \tilde{x}^{(q)}_{k-1}), C_j \tilde{y}_{k-1} \rangle|) + E(|\langle C_\ell \tilde{y}^{(q)}_{k-1}, C_j B(\tilde{x}_{k-1} - \tilde{x}^{(q)}_{k-1}) \rangle|) \big)$$

$$\leq \frac{1}{n} \sum_{k=1}^{n} \big(\|C_\ell B\|_{\mathcal{L}(H,K)} \, E(\|\tilde{x}_{k-1} - \tilde{x}^{(q)}_{k-1}\|_H \cdot \|C_j \tilde{y}_{k-1}\|) + \|C_j B\|_{\mathcal{L}(H,K)} \cdot E(\|C_\ell \tilde{y}^{(q)}_{k-1}\| \cdot \|\tilde{x}_{k-1} - \tilde{x}^{(q)}_{k-1}\|_H) \big)$$

$$= \|C_\ell B\|_{\mathcal{L}(H,K)} \cdot E(\|\tilde{x}_o - \tilde{x}_o^{(q)}\|_H \cdot \|C_j \tilde{y}_o\|) + \|C_j B\|_{\mathcal{L}(H,K)} \cdot E(\|C_\ell \tilde{y}_o^{(q)}\| \cdot \|\tilde{x}_o - \tilde{x}_o^{(q)}\|_H)$$

$$\leq \|C_\ell\|_{\mathcal{L}(K)} \cdot \|C_j\|_{\mathcal{L}(K)} \cdot \|B\|_{\mathcal{L}(H,K)} \cdot E(\|\tilde{x}_o - \tilde{x}_o^{(q)}\|_H \cdot (\|\tilde{y}_o\| + \|\tilde{y}_o^{(q)}\|))$$

$$\leq \text{const.} \cdot [E(\|\tilde{x}_o - \tilde{x}_o^{(q)}\|_H^2) \cdot E((\|\tilde{y}_o\| + \|\tilde{y}_o^{(q)}\|)^2)]^{1/2}.$$

Aufgrund der Beziehung $\tilde{x}_o - \tilde{x}_o^{(q)} = \sum_{\ell=q+1}^{\infty} A(\underline{\theta})\, v_{-\ell}$ erhält man

$$E(\|\tilde{x}_o - \tilde{x}_o^{(q)}\|^2) \leq \|A(\underline{\theta})\|_{\mathcal{L}(H)}^{2q+2} \cdot E(\|\tilde{x}_o\|^2).$$

$E(\|\tilde{y}_o\|)$ und $E(\|\tilde{y}_o^{(q)}\|)$ kann man ebenfalls durch eine (von q unabhängige) endliche Konstante abschätzen (vgl. Beweisteil a)), sodaß man zusammenfassen kann:

$$E(|\frac{1}{n} \sum_{k=1}^{n} \tilde{\beta}_{\ell,j}(k) - \frac{1}{n} \sum_{k=1}^{n} \tilde{\beta}_{\ell,j}^{(q)}(k)|) \leq d \cdot \|A(\underline{\theta})\|_{\mathcal{L}(H)}^{q+1} \quad \text{für alle } n,q \in \mathbb{N},\ 0 \leq d < \infty.$$

ii) Der Prozeß $(\tilde{\beta}_{\ell,j}^{(q)}(k))_{k\in\mathbb{N}} \equiv (\langle C_\ell \tilde{y}_{k-1}^{(q)},\ C_j \tilde{y}_{k-1}^{(q)} \rangle)_{k\in\mathbb{N}}$ ist ein φ-mischender Prozeß (mit einer i.allg. modifizierten Folge $(\varphi_k)_{k\in\mathbb{N}}$), da $\tilde{y}_k^{(q)}$ aus einem Block der Länge q+1 der Störvariablen $(v_\ell, e_\ell)_{\ell\in Z}$ aufgebaut ist. Insbesondere ist $(\tilde{\beta}_{\ell,j}^{(q)}(k))_{k\in\mathbb{N}}$ ein ergodischer Prozeß. Demzufolge besagt der Ergodensatz von Birkhoff, daß für alle $q \in \mathbb{N}$

$$\frac{1}{n} \sum_{k=1}^{n} \tilde{\beta}_{\ell,j}^{(q)}(k) \xrightarrow[n\to\infty]{} E(\langle C_\ell \tilde{y}_o^{(q)},\ C_j \tilde{y}_o^{(q)} \rangle) \quad \text{P-fast sicher.}$$

iii) Der dritte Term der Aufspaltung wird analog zum Vorgehen unter Punkt (i) abgeschätzt:

$$|E(\langle C_\ell \tilde{y}_o^{(q)},\ C_j \tilde{y}_o^{(q)} \rangle) - E(\langle C_\ell \tilde{y}_o, C_j \tilde{y}_o \rangle)|$$

$$\leq E(|\langle C_\ell (\tilde{y}_o - \tilde{y}_o^{(q)}),\ C_j \tilde{y}_o \rangle + \langle C_\ell \tilde{y}_o^{(q)},\ C_j (\tilde{y}_o - \tilde{y}_o^{(q)}) \rangle|) \leq d \cdot \|A(\underline{\theta})\|_{\mathcal{L}(H)}^{q+1}.$$

d) Die zu beweisende Konvergenzaussage (*), Teil b), folgt nun aus dem allgemeingültigen

Hilfssatz:

$(\beta_n)_{n\in\mathbb{N}}$ sei eine Folge von reellen Zufallsvariablen mit folgender Aufspaltung:

$$\beta_n = \alpha_n(q) + \gamma_n(q) \text{ für alle } n,q\in\mathbb{N}.$$

Es gelte: $\lim\limits_{q\to\infty} \sup\limits_{n\in\mathbb{N}} E(|\alpha_n(q)|) = 0$ und $\gamma_n(q) \xrightarrow[n\to\infty]{} 0$ P-stochastisch für alle $q\in\mathbb{N}$.

Dann konvergiert β_n P-stochastisch gegen 0 mit $n \to \infty$.

Beweis:

Seien $\varepsilon > 0$, $\delta > 0$ vorgegeben;

es ist $P(|\beta_n| > \varepsilon) \leq P(|\alpha_n(q)| > \frac{\varepsilon}{2}) + P(|\gamma_n(q)| > \frac{\varepsilon}{2})$

$\leq \frac{2}{\varepsilon} \sup\limits_{n\in\mathbb{N}} (E(|\alpha_n(q)|)) + P(|\gamma_n(q)| > \frac{\varepsilon}{2})$. Damit folgt die Behauptung. □

Wir erweitern nun die Aussage des obigen Lemmas auf den Fall, daß die Startvariable x_o nicht stationär ist.

Lemma II.3.2:

Für alle $\ell,j=1,\ldots,m$ konvergiert $\frac{1}{n}\gamma_{\ell,j}(n)$ P-fast sicher gegen

$E(<C_\ell y_o, C_j B\tilde{x}_o>)$ mit $n \to \infty$.

Beweis:

In Anbetracht von Lemma II.3.1 und den Beziehungen

$$\frac{1}{n}\gamma_{\ell,j}(n) = \left(\frac{1}{n}\cdot \sum_{k=1}^{n} <C_\ell y_{k-1}, C_j y_{k-1}>\right) - E(<C_\ell \tilde{y}_o, C_j e_o>),$$

$$E(<C_\ell \tilde{y}_o, C_j \tilde{y}_o>) - E(<C_\ell y_o, C_j e_o>) = E(<C_\ell \tilde{y}_o, C_j B\tilde{x}_o>)$$

genügt es zu zeigen, daß

$$z_n := \frac{1}{n}\sum_{k=1}^{n} <C_\ell y_{k-1}, C_j y_{k-1}> - \frac{1}{n}\sum_{k=1}^{n} <C_\ell \tilde{y}_{k-1}, C_j \tilde{y}_{k-1}> \xrightarrow[n\to\infty]{} 0 \text{ P-fast sicher.}$$

Per Definition der Größen y_k, y_K und weil $x_{k-1}-\tilde{x}_{k-1} = A(\underline{\theta})^{k-1}(x_0-\tilde{x}_0)$ ist, erhält man

$$z_n = \frac{1}{n}\sum_{k=1}^{n} (<C_\ell Bx_{k-1}, C_j Bx_{k-1}> - <C_\ell B\tilde{x}_{k-1}, C_j B\tilde{x}_{k-1}>)$$

$$+ \frac{1}{n}\sum_{k=1}^{n} <C_\ell BA(\underline{\theta})^{k-1}(x_0-\tilde{x}_0),\ C_j e_{k-1}>$$

$$+ \frac{1}{n}\sum_{k=1}^{n} <C_\ell e_{k-1},\ C_j BA(\underline{\theta})^{k-1}(x_0-\tilde{x}_0)>.$$

Der zweite (und damit auch der dritte) Term konvergiert P-fast sicher gegen 0, weil

$$\sum_{k=1}^{\infty} E(|<C_\ell BA(\underline{\theta})^{k-1}(x_0-\tilde{x}_0),\ C_j e_{k-1}>|)$$

$$\leq \sum_{k=1}^{\infty} E(\|C_\ell BA(\underline{\theta})^{k-1}\|_{\mathcal{L}(H,K)} \|x_0-\tilde{x}_0\|_H \cdot \|C_j e_{k-1}\|)$$

$$\leq \|C_\ell B\|_{\mathcal{L}(H,K)} \cdot \sum_{k=1}^{\infty} \|A(\underline{\theta})\|^{k-1}_{\mathcal{L}(H)} \cdot [E(\|x_0-\tilde{x}_0\|_H^2) \cdot E(\|C_j e_{k-1}\|^2)]^{1/2}$$

$< \infty$, da $E(\|x_0\|_H^2) < \infty$; d.h. $<C_\ell BA(\underline{\theta})^{k-1}(x_0-\tilde{x}_0),\ C_j e_{k-1}> \xrightarrow[k\to\infty]{} 0$ P-fast sicher.

Erst recht konvergiert das arithmetische Mittel gegen 0 P-fast sicher.

Der erste Term läßt sich umschreiben zu

$$\frac{1}{n}\sum_{k=1}^{n} <C_\ell BA(\underline{\theta})^{k-1}(x_0-\tilde{x}_0), C_j BA(\underline{\theta})^{k-1}(x_0-\tilde{x}_0)>$$

$$+ \frac{1}{n}\sum_{k=1}^{n} <C_\ell BA(\underline{\theta})(x_0-\tilde{x}_0),\ C_j B\tilde{x}_{k-1}>$$

$$+ \frac{1}{n}\sum_{k=1}^{n} <C_\ell B\tilde{x}_{k-1}, C_j BA(\underline{\theta})^{k-1}(x_0-\tilde{x}_0)>.$$

Der erste Summand konvergiert offensichtlich gegen 0. Für die Abschätzung des zweiten (bzw. dritten) Summanden benötigt man die Höldersche Ungleichung:

$$\sum_{k=1}^{\infty} E(|\langle C_\ell BA(\underline{\theta})^{k-1}(x_o-\tilde{x}_o), C_j B\tilde{x}_{k-1}\rangle|)$$

$$\leq \text{const.} \cdot \sum_{k=1}^{\infty} \|A(\underline{\theta})\|^{k-1}_{\mathcal{L}(H)} \cdot E(\|x_o-\tilde{x}_o\|_H \cdot \|\tilde{x}_{k-1}\|_H)$$

$$\leq \text{const.} \cdot \sum_{k=1}^{\infty} \|A(\underline{\theta})\|^{k-1}_{\mathcal{L}(H)} \cdot [E(\|x_o-\tilde{x}_o\|^2_H) \cdot E(\|\tilde{x}_{k-1}\|^2_H)]^{1/2} < \infty. \qquad \square$$

Lemma II.3.3:

Für $\ell = 1,\ldots,m$ gilt:

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \langle C_\ell \tilde{y}_{k-1}, Bv_k+e_k-C(\underline{\theta})e_{k-1}\rangle = E(\langle C_\ell \tilde{y}_o, Bv_1+e_1-C(\underline{\theta})e_o\rangle) \text{ P-fast sicher.}$$

Beweis:

Der Beweis läuft nach dem gleichen Argumentationsmuster ab wie derjenige zu Lemma II.3.1 und wird dementsprechend knapp dargestellt.

a) Für den strikt stationären Prozeß $\tilde{\beta}_\ell(k) := \langle C_\ell \tilde{y}_{k-1}, Bv_k+e_k-C(\underline{\theta})e_{k-1}\rangle$, $k\in\mathbb{N}$,

existiert $\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \tilde{\beta}_\ell(k)$ P-fast sicher, sodaß es zu zeigen genügt:

$$(*)\quad \frac{1}{n} \sum_{k=1}^{n} \tilde{\beta}_\ell(k) \xrightarrow[n\to\infty]{} E(\langle C_\ell \tilde{y}_o, Bv_1+e_1-C(\underline{\theta})e_o\rangle) \quad \text{P-stochastisch.}$$

b) Sei $\tilde{\beta}^{(q)}_\ell(k) := \langle C_\ell \tilde{y}^{(q)}_{k-1}, Bv_k+e_k-C(\underline{\theta})e_{k-1}\rangle$; dann spaltet man auf:

$$\frac{1}{n} \sum_{k=1}^{n} \tilde{\beta}_\ell(k) - E(\langle C_\ell \tilde{y}_o, Bv_1+e_1-C(\underline{\theta})e_o\rangle)$$

$$= [\frac{1}{n} \sum_{k=1}^{n} (\tilde{\beta}_\ell(k)-\tilde{\beta}^{(q)}_\ell(k))] + [\frac{1}{n} \sum_{k=1}^{n} \tilde{\beta}^{(q)}_\ell(k) - E(\langle C_\ell \tilde{y}^{(q)}_o, Bv_1+e_1-C(\underline{\theta})e_o\rangle)]$$

$$+ [E(\langle C_\ell \tilde{y}^{(q)}_o, Bv_1+e_1-C(\underline{\theta})e_o\rangle) - E(\langle C_\ell \tilde{y}_o, Bv_1+e_1-C(\underline{\theta})e_o\rangle)].$$

Weiterhin kann man abschätzen:

i) $E(|\frac{1}{n}\sum_{k=1}^{n}(\tilde{\beta}_\ell(k) - \tilde{\beta}_\ell^{(q)}(k))|) \leq$

$\frac{1}{n}\sum_{k=1}^{n} E(\|C_\ell(\tilde{y}_{k-1}-\tilde{y}_{k-1}^{(q)})\| \cdot \|Bv_k+e_k-C(\underline{\theta})e_{k-1}\|)$

$\leq \|C_\ell\|_{\mathfrak{L}(K)} \cdot [E(\|Bv_1+e_1-C(\underline{\theta})e_o\|^2)]^{1/2} \cdot [E(\|\tilde{y}_o-\tilde{y}_o^{(q)}\|^2)]^{1/2}$

$\leq$ const. $\cdot \|A(\underline{\theta})\|^{q+1}_{\mathfrak{L}(H)}$ für alle $n,q\in\mathbb{N}$.

ii) Da $(\tilde{\beta}_\ell^{(q)}(k))_{k\in\mathbb{N}}$ ein φ-mischender, also auch ergodischer Prozeß ist, gilt für alle $q\in\mathbb{N}$:

$\frac{1}{n}\sum_{k=1}^{n}\tilde{\beta}_\ell^{(q)}(k) \xrightarrow[n\to\infty]{} E(\langle C_\ell \tilde{y}_o^{(q)}, Bv_1+e_1-C(\underline{\theta})e_o\rangle)$ P-fast sicher.

iii) Schließlich ist

$|E(\langle C_\ell\tilde{y}_o^{(q)}, Bv_1+e_1-C(\underline{\theta})e_o\rangle) - E(\langle C_\ell\tilde{y}_o, Bv_1+e_1-C(\underline{\theta})e_o\rangle)|$

$\leq E(|\langle C_\ell(\tilde{y}_o-\tilde{y}_o^{(q)}), Bv_1+e_1-C(\underline{\theta})e_o\rangle|) \leq$ const. $\|A(\underline{\theta})\|^{q+1}_{\mathfrak{L}(H)}$

(vgl. Teil i)).

c) Die Aussage (*) folgt nun aus Teil b) und dem Hilfssatz aus Beweisteil d) zu Lemma II.3.1. □

Lemma II.3.4:

$\lim_{n\to\infty}\frac{1}{n}(\underline{\eta}(n) - \Gamma(n)\underline{\theta}) = \underline{0}$ P-fast sicher.

Beweis: Aufgrund der Definition der Größen $\underline{\eta}(n)$ und $\Gamma(n)$ (vgl. § II.2) und der Beziehung $y_k - C(\underline{\theta})y_{k-1} = Bv_k + e_k - C(\underline{\theta})e_{k-1}$ erhalten wir

$\frac{1}{n}(\underline{\eta}(n)-\Gamma(n)\underline{\theta})_\ell = \frac{1}{n}(\eta_\ell(n) - \sum_{j=1}^{m}\gamma_{\ell,j}(n)\theta_j) =$

$\frac{1}{n}\cdot(\sum_{k=1}^{n}[\langle C_\ell y_{k-1}, y_k - C_o y_{k-1}\rangle - E(\langle C_\ell\tilde{y}_o, Bv_1+e_1-C_o e_o\rangle)]$

$-\sum_{j=1}^{m}\theta_j\cdot\sum_{k=1}^{n}[\langle C_\ell y_{k-1}, C_j y_{k-1}\rangle - E(\langle C_\ell\tilde{y}_o, C_j e_o\rangle)])$

$$= \frac{1}{n} \cdot \sum_{k=1}^{n} \left([<C_\ell y_{k-1}, y_k - C_o y_{k-1}> - E(<C_\ell \tilde{y}_o, Bv_1 + e_1 - C_o e_o>)] \right.$$

$$\left. - [<C_\ell y_{k-1}, (C(\underline{\theta}) - C_o) y_{k-1}> - E(<C_\ell \tilde{y}_o, (C(\underline{\theta}) - C_o) e_o>)] \right)$$

$$= \frac{1}{n} \cdot \sum_{k=1}^{n} [<C_\ell y_{k-1}, y_k - C(\underline{\theta}) y_{k-1}> - E(<C_\ell \tilde{y}_o, Bv_1 + e_1 - C(\underline{\theta}) e_o>)]$$

$$= \left(\frac{1}{n} \cdot \sum_{k=1}^{n} <C_\ell y_{k-1}, Bv_k + e_k - C(\underline{\theta}) e_{k-1}> \right) - E(<C_\ell \tilde{y}_o, Bv_1 + e_1 - C(\underline{\theta}) e_o>).$$

In Anbetracht von Lemma II.3.3 bleibt deshalb zu zeigen:

$$z_n := \frac{1}{n} \sum_{k=1}^{n} <C_\ell y_{k-1}, Bv_k + e_k - C(\underline{\theta}) e_{k-1}> - \frac{1}{n} \sum_{k=1}^{n} <C_\ell \tilde{y}_{k-1}, Bv_k + e_k - C(\underline{\theta}) e_{k-1}> \xrightarrow[n \to \infty]{} 0 \text{ P-fast sicher.}$$

Es ist $z_n = \frac{1}{n} \sum_{k=1}^{n} <C_\ell BA(\underline{\theta})^{k-1} (x_o - \tilde{x}_o), Bv_k + e_k - C(\underline{\theta}) e_{k-1}>$.

Ferner läßt sich abschätzen:

$$\sum_{k=1}^{\infty} E(| <C_\ell BA(\underline{\theta})^{k-1} (x_o - \tilde{x}_o), Bv_k + e_k - C(\underline{\theta}) e_{k-1}> |)$$

$$\leq \| C_\ell B \|_{\mathcal{L}(H,K)} \cdot \sum_{k=1}^{\infty} \| A(\underline{\theta}) \|_{\mathcal{L}(H)}^{k-1} \cdot [E(\| x_o - \tilde{x}_o \|_H^2) \cdot E(\| Bv_k + e_k - C(\underline{\theta}) e_{k-1} \|^2)]^{1/2} < \infty;$$

also $<C_\ell BA(\underline{\theta})^{k-1} (x_o - \tilde{x}_o), Bv_k + e_k - C(\theta) e_{k-1}> \xrightarrow[k \to \infty]{} 0$ P-fast sicher. Demnach konvergiert das arithmetische Mittel, nämlich z_n, auch P-fast sicher gegen 0 mit $n \to \infty$. □

Wir können nun das Hauptergebnis dieses Abschnitts formulieren, das das Parameterschätzproblem bez. des im Abschnitt II.1 (durch die Gleichungen (I), (II)) beschriebenen Systems betrifft. Im folgenden Satz benötigen wir die Matrix Γ, die definiert ist als $(E(<C_\ell \tilde{y}_o, C_j B \tilde{x}_o>))_{\substack{\ell=1,\ldots,m \\ j=1,\ldots,m}}$, wobei die Operatoren C_ℓ in der Beobachtbarkeitsbedingung (O) (II.1) und die Variablen $\tilde{y}_o$, $\tilde{x}_o$ (stationäre Varianten von y_k bzw. x_k) anfangs dieses Abschnitts erklärt worden sind. B ist der in der Beobachtungsgleichung (II) auftretende Operator aus $\mathcal{L}(H,K)$.

Satz II.3.1 (Starke Konsistenz des Schätzers $\hat{\underline{\theta}}_n$)

Falls die Matrix Γ invertierbar ist, ist der im Abschnitt II.2 konstruierte Least-Square-Schätzer $\hat{\underline{\theta}}_n$ eindeutig bestimmt für alle $n \geq n_o(\omega)$. In diesem Fall konvergiert $\hat{\underline{\theta}}_n$ P-fast sicher gegen den (wahren) Parameter $\underline{\theta}$ mit $n \to \infty$.

Beweis:

Gemäß Lemma II.3.4 konvergiert $\frac{1}{n}\Gamma(n)\,(\hat{\underline{\theta}}_n - \underline{\theta}) = \frac{1}{n}\,(\underline{\eta}(n) - \Gamma(n)\underline{\theta})$ gegen $0 \in \mathbb{R}^m$ P-fast sicher. Weiterhin konvergiert $\frac{1}{n}\Gamma(n)$ P-fast sicher gegen Γ, vgl. Lemma II.3.2. Aufgrund der vorausgesetzten Invertierbarkeit von Γ und der Tatsache, daß $\frac{1}{n}\Gamma(n)$ im endlich-dimensionalen Raum $\mathbb{R}^m$ operiert, erhält man die Existenz von $(\frac{1}{n}\Gamma(n,\omega))^{-1}$ für $n \geq n_o(\omega)$, und auch die Beschränktheit der Folge $(\|(\frac{1}{n}\Gamma(n,\omega))^{-1}\|_{\mathcal{L}(\mathbb{R}^m)})_{n \geq n_o(\omega)}$ für P-fast alle $\omega \in \Omega$. Letzteres ergibt zusammen mit den oben festgestellten Konvergenzen die Behauptung des Satzes. □

§ II.4: Asymptotische Normalität des Kleinste-Quadrate-Schätzers

In diesem Abschnitt leiten wir die asymptotische Normalität des Schätzers $\hat{\underline{\theta}}_n$ her, d.h. wir zeigen, daß $\sqrt{n}\,(\hat{\underline{\theta}}_n - \underline{\theta})$ mit $n \to \infty$ schwach gegen eine m-dimensionale Normalverteilung konvergiert. Der Beweis dieser Tatsache stützt sich im wesentlichen auf einen Zentralen Grenzwertsatz für Funktionen von φ-mischenden Zufallsgrößen in der Form des Theorems 21.1, [BI, '68], S. 184. Wir führen zuvor einige Bezeichnungen ein. Für alle $\underline{r} \in \mathbb{R}^m$ definiere man den Operator $D(\underline{r}) := \sum_{\ell=1}^{m} r_\ell C_\ell \in \mathcal{L}(K)$. Sei

$$z_k^{(q)} := z_k^{(q)}(\underline{r}) := \langle D(\underline{r})\tilde{y}_{k-1}^{(q)},\ Bv_k + e_k - C(\underline{\theta})e_{k-1}\rangle - E(\langle D(\underline{r})\tilde{y}_o, Bv_1 + e_1 - C(\underline{\theta})e_o\rangle)$$

und $z_k := z_k(\underline{r}) := \langle D(\underline{r})\tilde{y}_{k-1},\ Bv_k + e_k - C(\underline{\theta})e_{k-1}\rangle - E(\langle D(\underline{r})\tilde{y}_o, Bv_1 + e_1 - C(\underline{\theta})e_o\rangle)$

für $q,k \in \mathbb{N}$ und $\underline{r} \in \mathbb{R}^m$. Man betrachte ferner die Partialsummen

$$\tilde{S}_n := \tilde{S}_n(\underline{\theta},\underline{r}) := \sum_{k=1}^{n} z_k(\underline{r}),\ n \in \mathbb{N},\ \underline{r} \in \mathbb{R}^m.$$

Für den gesamten Abschnitt II.4 verlangen wir die (im Vergleich mit § II.3) schärferen Momentenbedingungen $E(\|v_o\|_H^4) < \infty$ und $E(\|e_o\|^4) < \infty$. Schließlich müssen wir noch folgende Konvergenzgeschwindigkeit für die Mischungskoeffizienten $(\varphi_n)_{n\in\mathbb{N}}$ festlegen:

$$\sum_{n=1}^{\infty} \varphi_n^{1/2} < \infty.$$

Dann gilt folgendes

Lemma II.4.1:

Die Reihe $\sigma^2 := \sigma^2(\underline{\theta},\underline{r}) := E(z_1(\underline{r})^2) + 2 \cdot \sum_{k=2}^{\infty} E(z_1(\underline{r}) \cdot z_k(\underline{r}))$

konvergiert absolut und es gilt $\sigma^2 = \lim_{q\to\infty} \sigma^2(q)$, wobei

$\sigma^2(q) := \sigma^2(q,\underline{\theta},\underline{r}) := E(z_1^{(q)}(\underline{r})^2) + 2 \cdot \sum_{k=2}^{\infty} E(z_1^{(q)}(\underline{r}) \cdot z_k^{(q)}(\underline{r}))$.

$\frac{1}{\sqrt{n}} \tilde{S}_n$ konvergiert schwach gegen die Normalverteilung $\mathcal{N}(0,\sigma^2)$ mit Mittelwert 0 und Varianz σ^2.

Beweis:

Es genügt, die Voraussetzungen des Theorems 21.1, [BI, '68], zu verifizieren. Dabei ist zu beachten, daß die φ-mischende Ausgangsfolge nicht reellwertig zu sein braucht (vgl. die Bemerkung in [BI, '68], S. 182). Ferner ist auch der Wert $\sigma^2 = 0$ zugelassen, wie die Bemerkung im Anschluß an den Beweis des Th. 21.1, S. 190, feststellt.

Es wird nun gezeigt, wie sich die Variablen z_k, $k \in \mathbb{N}$, als Funktionen der φ-mischenden Folge $(v_k,e_k)_{k \in Z}$ schreiben lassen und daß sich eine hinreichende Approximationsgüte (im quadratischen Mittel) durch die Variablen $z_k^{(q)}$ ergibt, die nur aus einem Block der Länge $q + 1$ (bzw. $2q + 1$) der Störvariablen gebildet werden.

Sei $c := E(\langle D(\underline{r})\tilde{y}_o, Bv_1 + e_1 - C(\underline{\theta})e_o\rangle)$ und die Abbildung $f: (H \times K)^Z \to \mathbb{R}$ wie folgt definiert ($(h_j,k_j) \in H \times K$ stets):

$f(\ldots, (h_{-2},k_{-2}), (h_{-1},k_{-1}), (h_o,k_o), (h_1,k_1), (h_2,k_2),\ldots) :=$

$\langle D(\underline{r})\,[B(\sum_{j=0}^{\infty} A(\underline{\theta})^j h_{-1-j}) + k_{-1}],\ Bh_o + k_o - C(\underline{\theta})k_{-1}\rangle - c,$

falls $\lim_{q\to\infty} \sum_{j=0}^{q} A(\underline{\theta})^j h_{-1-j}$ existiert; $f(\ldots) = 0$ sonst.

Für $q \in \mathbb{N}$ erkläre man $f_q: (H \times K)^{2q+1} \to \mathbb{R}$ durch

$f_q((h_{-q},k_{-q}),\ldots,(h_{-1},k_{-1}),(h_o,k_o),(h_1,k_1),\ldots,(h_q,k_q)) :=$

$\langle D(\underline{r})\,[B(\sum_{j=0}^{q-1} A(\underline{\theta})^j h_{-1-j}) + k_{-1}],\ Bh_o + k_o - C(\underline{\theta})k_{-1}\rangle - c.$

Lemma II.4.3:

Der Limes $\sigma^2_{\ell,j} := \sigma^2_{\ell,j}(\underline{\theta}) := \lim_{n\to\infty} \sigma^2_n(\underline{\theta};\ell,j)$ existiert für alle $\ell,j = 1,\ldots,m$ und $\frac{1}{\sqrt{n}}(\underline{n}(n) - \Gamma(n)\underline{\theta})$ konvergiert schwach gegen die m-dimensionale Normalverteilung $\mathcal{N}(\underline{0},(\sigma^2_{\ell,j})_{\ell,j=1,\ldots,m})$.

Beweis:

a) Zunächst zeigen wir, daß $\sigma^2_{\ell,j} = \lim_{n\to\infty} \sigma^2_n(\underline{\theta},\ell,j)$ existiert für alle $\ell,j = 1,\ldots,m$.

Wegen $D(\underline{r}) = \sum_{\ell=1}^{m} r_\ell C_\ell$ gilt $z_k(\underline{r}) = \sum_{\ell=1}^{m} r_\ell z_k(\underline{e}_\ell)$

und deshalb $\sigma^2(\underline{\theta},\underline{r}) = \lim_{n\to\infty} [E(z_1(\underline{r})^2) + 2 \cdot \sum_{k=2}^{n} E(z_1(\underline{r}) \cdot z_k(\underline{r}))]$

$= \lim_{n\to\infty} [\sum_{\ell,j=1}^{m} r_\ell r_j \cdot [E(z_1(\underline{e}_\ell) \cdot z_1(\underline{e}_j)) + 2 \cdot \sum_{k=2}^{n} E(z_1(\underline{e}_\ell) \cdot z_k(\underline{e}_j))]]$

$= \lim_{n\to\infty} \sum_{\ell,j=1}^{m} r_\ell r_j \sigma^2_n(\underline{\theta};\ell,j)$.

Aufgrund der folgenden Hilfsaussage folgt dann die Behauptung:

Hilfssatz: Für alle $n \in \mathbb{N}$ sei $A(n) = (\alpha_{\ell,j}(n))_{\substack{\ell=1,\ldots,m \\ j=1,\ldots,m}}$ eine symmetrische m×m-Matrix derart, daß für alle $\underline{r} \in \mathbb{R}^m$ $(\underline{r},A(n)\underline{r})_m = \sum_{\ell,j=1}^{m} r_\ell r_j \alpha_{\ell,j}(n)$ konvergiert mit $n \to \infty$. Dann konvergiert auch $\alpha_{\ell,j}(n)$ mit $n \to \infty$ für alle $\ell,j = 1,\ldots,m$.

Beweis:

Aufgrund der Symmetrie der Matrix $A(n)$ erhält man $\alpha_{\ell,j}(n) = (\underline{e}_\ell, A(n)\underline{e}_j)_m =$

$$\frac{1}{2}\left[(\underline{e}_\ell + \underline{e}_j, A(n)(\underline{e}_\ell + \underline{e}_j))_m - (\underline{e}_\ell, A(n)\underline{e}_\ell)_m - (\underline{e}_j, A(n)\underline{e}_j)_m\right],$$

und damit die Behauptung.—

b) Um die m-dimensionale schwache Konvergenz zu verifizieren, reicht es angesichts des "Cramér-Wold-Device" (s. z.B. [GÄ/ST, '77]) zu zeigen, daß für alle $r = (r_1, \ldots, r_m) \in \mathbb{R}^m$ gilt

$$\sum_{\ell=1}^{m} r_\ell \frac{1}{\sqrt{n}} (\underline{\eta}(n) - \Gamma(n)\underline{\theta})_\ell \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0, \sum_{\ell,j=1}^{m} r_\ell r_j \sigma^2_{\ell,j}).$$

Einerseits ist $\sigma^2(\underline{\theta},\underline{r}) = \lim_{n\to\infty} \sum_{\ell,j=1}^{m} r_\ell r_j \sigma_n^2(\underline{\theta},\ell,j) = \sum_{\ell,j=1}^{m} r_\ell r_j \sigma^2_{\ell,j}$.

Andererseits erhalten wir (vgl. Beweis von Lemma II.3.4):

$$\sum_{\ell=1}^{m} r_\ell \frac{1}{\sqrt{n}} (\underline{\eta}(n) - \Gamma(n)\underline{\theta})_\ell =$$

$$\frac{1}{\sqrt{n}} \sum_{\ell=1}^{m} r_\ell \sum_{k=1}^{n} \left[<C_\ell y_{k-1}, y_k - C(\underline{\theta}) y_{k-1}> - E(<C_\ell \tilde{y}_o, Bv_1 + e_1 - C(\underline{\theta})e_o>)\right]$$

$$= \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \left[<D(\underline{r})y_{k-1}, y_k - C(\underline{\theta})y_{k-1}> - E(<D(\underline{r})\tilde{y}_o, Bv_1 + e_1 - C(\underline{\theta})e_o>)\right]$$

$$= \frac{1}{\sqrt{n}} S_n(\underline{\theta},\underline{r}).$$ Somit folgt die Behauptung aus Lemma II.4.2.

□

Wir geben nun das Hauptergebnis dieses Abschnitts an.

Satz II.4.1 (Asymptotische Normalität des Least-Square-Schätzers $\hat{\underline{\theta}}_n$):
Unter denselben Annahmen wie im Satz II.3.1 konvergiert $\sqrt{n}\,(\hat{\underline{\theta}}_n - \underline{\theta})$ schwach gegen die m-dimensionale Normalverteilung $\mathcal{N}(\underline{0},\ \Gamma^{-1}\Sigma(\Gamma^{-1})^*)$ mit $n \to \infty$.
Dabei ist Σ die Matrix bestehend aus den Elementen

$$\sigma^2_{\ell,j} = E(z_1(\underline{e}_\ell)\cdot z_1(\underline{e}_j)) + 2\cdot \sum_{k=2}^{\infty} E(z_1(\underline{e}_\ell)\cdot z_k(\underline{e}_j)),$$

wobei die Variablen $z_i(\underline{r})$, $\underline{r}\in \mathbb{R}^m$, zu Beginn dieses Abschnitts definiert sind, und die $\underline{e}_i$ Einheitsvektoren im $\mathbb{R}^m$ sind.
Im Falle der Invertierbarkeit der Matrix Σ kann die Konvergenzaussage noch verschärft werden:
Sei Q_n die Verteilung von $\sqrt{n}\,(\hat{\underline{\theta}}_n - \underline{\theta})$ und $\mathcal{C}$ das System aller Borel-meßbaren, konvexen Mengen im $\mathbb{R}^m$. Dann gilt: $\sup\{|Q_n(C) - \mathcal{N}(\underline{0},\Gamma^{-1}\Sigma(\Gamma^{-1})^*)(C)|: C\in\mathcal{C}\} \xrightarrow[n\to\infty]{} 0$.
Insbesondere konvergiert also die zu $\sqrt{n}\,(\hat{\underline{\theta}}_n - \underline{\theta})$ gehörige Folge der Verteilungsfunktionen gleichmäßig gegen diejenige der $\mathcal{N}(\underline{0},\Gamma^{-1}\Sigma(\Gamma^{-1})^*)$-Verteilung.

Beweis:

Es gilt (vgl. Beweis zu Satz II.3.1) für $n \geq n_0(\omega)$:

$$\sqrt{n}\,(\hat{\underline{\theta}}_n - \underline{\theta}) = \left(\frac{1}{n}\Gamma(n)\right)^{-1} \frac{1}{\sqrt{n}}\,(\underline{n}(n) - \Gamma(n)\underline{\theta}).$$

Gemäß Lemma II.3.2 konvergiert $\left(\frac{1}{n}\Gamma(n)\right)^{-1}$ P-fast sicher gegen $\Gamma^{-1} = \text{const.}$ (wobei man $\left(\frac{1}{n}\Gamma(n)\right)^{-1} := 0$ setzen kann, falls $\Gamma(n)$ nicht invertierbar ist).
Andererseits konvergiert $\frac{1}{\sqrt{n}}\,(\underline{n}(n) - \Gamma(n)\underline{\theta})$ schwach gegen $\mathcal{N}(\underline{0},\Sigma)$.
Sei $\underline{Z}$ ein $\mathcal{N}(\underline{0},\Sigma)$-verteilter Zufallsvektor. Aufgrund des Korollars 8.6.5 [GÄ/ST, '77], S. 353, wissen wir, daß

$$\sqrt{n}\,(\hat{\underline{\theta}}_n - \underline{\theta}) \xrightarrow[n\to\infty]{\mathcal{L}} \Gamma^{-1}\underline{Z} \overset{\mathcal{L}}{=} \mathcal{N}(\underline{0},\Gamma^{-1}\Sigma(\Gamma^{-1})^*).$$

Dazu beachte man, daß die Abbildung $h: \mathcal{L}(\mathbb{R}^m) \times \mathbb{R}^m \longrightarrow \mathbb{R}^m$, $h(M,z) := Mz$ stetig ist. –

Für invertierbares Σ ist $\mathcal{N}(\underline{0}, \Gamma^{-1}\Sigma(\Gamma^{-1})^*)$ eine nicht-ausgeartete, m-dimensionale Normalverteilung, also absolut-stetig bezüglich des Lebesgueschen Maßes im $\mathbb{R}^m$. Die verschärfte Konvergenzaussage ergibt sich somit unmittelbar aus dem Theorem 4.2, [RA, '62], S. 665.

□

§ II.5: Spezialisierung des dynamischen Modells auf den Fall unabhängiger Störvariabler

Wir behandeln hier einen Spezialfall des in den vorherigen Abschnitten betrachteten Modells. Und zwar nehmen wir an, daß die Systemstörungen $(v_k)_{k\in Z}$ und die Meßfehler $(e_k)_{k\in Z}$ jeweils eine Folge von unabhängigen, identisch verteilten Variablen sind, die außerdem als Folgen voneinander unabhängig sind. In dieser Situation lassen sich viele Formeln expliziter angeben, zum Teil in einer Gestalt, in der (seitens der stochastischen Struktur) nur die Kovarianzoperatoren der Störungen S_v bzw. S_e auftauchen. U.a. kann man plausible, hinreichende Kriterien für die bedeutsame Voraussetzung der Invertierbarkeit der Matrix Γ (vgl. die Sätze II.3.1 und II.4.1) angeben und diese im statistischen Zusammenhang erläutern. Außerdem wird man hier feststellen, daß man für die asymptotische Normalität des Schätzers $\hat{\underline{\theta}}_n$ zwar wiederum die Momentenbedingung $E(\|e_o\|^4) < \infty$ braucht, aber anstelle von $E(\|v_o\|_H^4) < \infty$ die Voraussetzung $E(\|v_o\|_H^2) < \infty$ genügt.–

Folgende (einfach zu verifizierende) Rechenregeln, die wir im weiteren ohne besondere Erwähnung häufig verwenden, seien ohne Beweis angegeben.

Lemma II.5.1:

Für zwei K-wertige, unabhängige Zufallselemente ξ und η, deren Erwartungswerte existieren, gilt:

(a) $A,B \in \mathfrak{L}(K)$, $E(\|\xi\|^2) < \infty \Rightarrow E(\langle A\xi, B\xi\rangle) = tr(AS_\xi B^*)$, wobei S_ξ den zur Verteilung $P\bullet\xi$ gehörigen Kovarianzoperator bezeichnet;

(b) $E(\langle\eta,\xi\rangle) = \langle E(\eta), E(\xi)\rangle$;

(c) $E(\|\eta\|^2) < \infty$ und $E(\|\xi\|^2) < \infty \Rightarrow E(\langle\eta,\xi\rangle^2) = tr(S_\eta S_\xi)$;

(d) $E(\eta) = E(\xi) = 0$, $E(\|\eta\|^2) < \infty$ und $E(\|\xi\|^2) < \infty \Rightarrow S_{\eta+\xi} = S_\eta + S_\xi$.

Außerdem gilt folgende Verallgemeinerung von (c):

(e) Seien $\eta\colon \Omega \to H$, $\xi\colon \Omega \to K$ stochastisch unabhängig mit existierenden zweiten Momenten, $A,C \in \mathcal{L}(H,K), B,D \in \mathcal{L}(K)$

$\Rightarrow E(\langle A\eta, B\xi\rangle \langle C\eta, D\xi\rangle) = \mathrm{tr}(DS_\xi B^* A S_\eta C^*)$.

Zunächst berechnen wir die Erwartungswerte, die in den Größen $\Gamma(n)$ und $\underline{\eta}(n)$, welche den Schätzer $\hat{\underline{\theta}}_n$ bestimmen, vorkommen (vgl. § II.2).

Es ist $E(\langle C_\ell \dot{y}_o, C_j e_o\rangle) = E(\langle C_\ell B x_o, C_j e_o\rangle) + E(\langle C_\ell e_o, C_j e_o\rangle)$

$= E(\langle C_\ell e_o, C_j e_o\rangle) = \mathrm{tr}(C_\ell S_e C_j^*)$ und

$E(\langle C_\ell \tilde{y}_o, Bv_1 + e_1 - C_o e_o\rangle) =$

$E(\langle C_\ell B\tilde{x}_o, Bv_1 + e_1 - C_o e_o\rangle) + E(\langle C_\ell e_o, Bv_1 + e_1 - C_o e_o\rangle) =$

$- E(\langle C_\ell e_o, C_o e_o\rangle) = - \mathrm{tr}(C_\ell S_e C_o^*)$, da $\tilde{x}_o$ unabhängig von $\sigma((e_k)_{k\in Z}, (v_k)_{k\in \mathbb{N}})$ ist.

Insbesondere sind diese Erwartungswerte nicht von dem unbekannten Parameter $\underline{\theta}$ abhängig.

Satz II.3.1 kann man nun umformulieren zu

Satz II.5.1:

Die Matrix $\Gamma := \Gamma(\underline{\theta}) := (\mathrm{tr}(BA_\ell F(\underline{\theta}) A_j^* B^*))_{\substack{\ell=1,\dots,m \\ j=1,\dots,m}}$ sei invertierbar, wobei

$F(\underline{\theta}) := \sum_{\kappa=0}^{\infty} A(\underline{\theta})^\kappa S_v (A(\underline{\theta})^*)^\kappa$ der zur Verteilung $P\bullet\tilde{x}_o$ gehörige Kovarianzoperator ist. Dann konvergiert der Least-Square Schätzer $\hat{\underline{\theta}}_n$ P-fast sicher gegen $\underline{\theta}$ mit $n \to \infty$.

Beweis:

Es ist $E(\langle C_\ell y_o, C_j B x_o\rangle) =$

$E(\langle C_\ell B\tilde{x}_o, C_j B\tilde{x}_o\rangle) + E(\langle C_\ell e_o, C_j B\tilde{x}_o\rangle) =$

$\mathrm{tr}(C_\ell B S_{\tilde{x}_o} (C_j B)^*) = \mathrm{tr}(BA_\ell F(\underline{\theta}) A_j^* B^*)$. □

Wir diskutieren nun Kriterien für die Invertierbarkeit von $\Gamma(\underline{\theta})$.

In der eindimensionalen Situation, d.h. $x_{k+1} = \theta a x_k + v_{k+1}, y_k = b x_k + e_k$, θ, a, b reell, gilt (falls $|a\theta| < 1$)

$$F(\theta) = \sum_{\kappa=0}^{\infty} (a\theta)^{2\ell}\sigma^2 = \frac{\sigma^2}{1-(a\theta)^2} \quad \text{mit } \sigma^2 = E(v_o^2),$$

$$\text{also ist } \Gamma(\theta) = a^2 b^2 F(\theta) = \frac{a^2 b^2 \sigma^2}{1-a^2\theta^2}.$$

Deshalb ist $\Gamma(\theta) \neq 0$ äquivalent damit, daß $a \neq 0$, $b \neq 0$ und $\sigma^2 > 0$. Die letzteren Bedingungen kann man einfach interpretieren (dabei sei $x_o = 0$): $\sigma^2 > 0$ ist gleichbedeutend mit $(x_k)_{k\in\mathbb{N}} \neq 0$; $a \neq 0$ bedeutet, daß θ überhaupt die Entwicklung der x_k beeinflußt; $b \neq 0$ ist äquivalent damit, daß die x_k "beobachtbar" sind mittels der y_k, d.h. hier, daß die Größen x_k in die Größen y_k eingehen.

Im mehrparametrigen, mehrdimensionalen Fall können ähnliche Kriterien für die Invertierbarkeit von $\Gamma(\underline{\theta})$ aufgestellt werden.

Korollar II.5.1:

Für $\underline{r} = (r_1,\ldots,r_m) \in \mathbb{R}^m$ sei der Operator $\tilde{A}(\underline{r}) \in \mathcal{L}(H)$ gegeben durch $\tilde{A}(\underline{r}) := \sum_{\ell=1}^{m} r_\ell A_\ell$. Der Schätzer $\hat{\underline{\theta}}_n$ ist stark konsistent, falls

(+) $P(v_1 \in \mathrm{Kern}(B\tilde{A}(\underline{r}))) < 1$ für alle $\underline{r} \in \mathbb{R}^m \setminus \{\underline{0}\}$.

Beweis:

Es bleibt nachzuweisen, daß $\Gamma(\underline{\theta})$ positiv-definit, also invertierbar ist. Nun kann man abschätzen:

$$(\underline{r}, \Gamma(\underline{\theta})\underline{r})_m = \sum_{\ell,j=1}^{m} r_\ell r_j \mathrm{tr}(BA_\ell F(\underline{\theta})(BA_j)^*)$$

$$= \mathrm{tr}(B\tilde{A}(\underline{r})F(\underline{\theta})(\tilde{A}(\underline{r}))^*)$$

$$= \mathrm{tr}(B\tilde{A}(\underline{r})S_v(B\tilde{A}(\underline{r}))^*) + \mathrm{tr}(B\tilde{A}(\underline{r})A(\underline{\theta})F(\underline{\theta})(B\tilde{A}(\underline{r})A(\underline{\theta}))^*)$$

$$\geq \mathrm{tr}(B\tilde{A}(\underline{r})S_v(B\tilde{A}(\underline{r}))^*), \text{ da}$$

$F(\underline{\theta}) = \sum_{\kappa=0}^{\infty} A(\underline{\theta})^\kappa S_v (A(\underline{\theta})^*)^\kappa = S_v + A(\underline{\theta})F(\underline{\theta})A(\underline{\theta})^*$ und $F(\underline{\theta})$ nichtnegativ-definit ist.

Aber $\mathrm{tr}(\tilde{BA}(\underline{r})S_v(\tilde{BA}(\underline{r}))^*) = E(\|\tilde{BA}(\underline{r})v_1\|^2) > 0$ für alle $\underline{r}\in\mathbb{R}^m\setminus\{\underline{0}\}$

wegen der Voraussetzung (+). □

Wir bemerken noch, daß in folgender Situation die Bedingung (+) aus Korollar II.5.1 erfüllt ist:

(i) $P(v_1 \in \mathrm{Kern}(BA_\ell)) < 1$ für alle $\ell = 1,\dots,m$

und

(ii) $(BA_\ell)(H)$ ist orthogonal zu $(BA_j)(H)$ für alle $\ell \neq j$, $\ell,j=1,\dots,m$.

Es ist nämlich $\mathrm{Kern}\,(\tilde{BA}(\underline{r})) = \{x\in H: \sum_{\ell=1}^m r_\ell BA_\ell x = 0\}$

$= \{x\in H: \langle \sum_{\ell=1}^m r_\ell BA_\ell x, \sum_{\ell=1}^m r_\ell BA_\ell x\rangle = 0\}$

$= \{x\in H: \sum_{\ell,j=1}^m r_\ell r_j \langle BA_\ell x, BA_j x\rangle = 0\}$

$\overset{(ii)}{=} \{x\in H: \sum_{\ell=1}^m r_\ell^2 \|BA_\ell x\|^2 = 0\}.$

Falls also $\underline{r}\neq 0$ ist, existiert $\ell_0 \in \{1,\dots,m\}$ mit $\mathrm{Kern}\,(\tilde{BA}(\underline{r})) \subset \mathrm{Kern}(\tilde{BA}_{\ell_0})$; damit gilt

$$P(v_1 \in \mathrm{Kern}(\tilde{BA}(\underline{r}))) \leq \max_{\ell=1,\dots,m} P(v_1 \in \mathrm{Kern}(BA_\ell)) < 1$$

für alle $\underline{r}\in\mathbb{R}^m\setminus\{0\}$, also Bedingung (+).–

Im einparametrigen Fall genügt die Bedingung $P(v_1 \in \mathrm{Kern}(BA_1)) < 1$, die eine plausible statistische Erläuterung erlaubt. Falls nämlich diese Voraussetzung nicht erfüllt ist, gilt $C_1 Bv_k = BA_1 v_k = 0$ P-fast sicher, und damit auch $BA_1^\ell v_k = 0$ P-fast sicher für alle $k,\ell \in \mathbb{N}$; denn

$BA_1^{\ell+1} v_k = BA_1 A_1^\ell v_k = C_1(BA_1^\ell v_k)$. Also gilt $Bx_k = Bv_k$ P-fast sicher für alle $k\in\mathbb{N}$ (unabhängig von θ), sodaß offensichtlich statistisches Schätzen von θ aufgrund der Beobachtungen $y_k = Bx_k + e_k$ unmöglich ist (der Einfachheit halber haben wir $A_0 = 0$ und $x_0 = 0$ angenommen).

Wir kommen nun zum Themenkreis der asymptotischen Normalität des Schätzers. Um explizite Formeln für die Kovarianzmatrix der Grenzverteilung angeben zu können, brauchen wir zunächst analytische Ausdrücke für Erwartungswerte der Form $E(z_l^{(q)} \cdot z_k^{(q)})$ für alle $k,q \in \mathbb{N}$. Dabei ist $z_k^{(q)} = z_k^{(q)}(\underline{r})$ genauso definiert wie zu Beginn des Abschnitts II.4; nämlich

$$z_k^{(q)}(\underline{r}) = \langle D(\underline{r})\tilde{y}_{k-1}^{(q)}, Bv_k + e_k - C(\underline{\theta})e_{k-1}\rangle - E(\langle D(\underline{r})\tilde{y}_o, Bv_l + e_l - C(\underline{\theta})e_o\rangle);$$

außerdem gilt hier

$$E(\langle D(\underline{r})\tilde{y}_o, Bv_l + e_l - C(\underline{\theta})e_o\rangle = -E(\langle D(\underline{r})e_o, C(\underline{\theta})e_o\rangle) = -\mathrm{tr}(D(\underline{r})S_e C(\underline{\theta})^*).$$

Lemma II.5.2:

Es sei $E(\|v_o\|_H^2) < \infty$ und $E(\|e_o\|^4) < \infty$. Dann gelten für alle $q \in \mathbb{N}$ folgende Formeln:

$$(i) E((z_l^{(q)})^2) = \sum_{j,\ell=1}^{m} r_j r_\ell \{\mathrm{tr}[C_j(B\sum_{\kappa=0}^{q} A(\underline{\theta})^\kappa S_v (A(\underline{\theta})^*)^\kappa)B^* + S_e)C_\ell^*(BS_vB^* + S_e)$$

$$+ C_j B(\sum_{\kappa=0}^{q} A(\theta)^\kappa S_v (A(\underline{\theta})^*)^\kappa)B^* C_\ell C(\underline{\theta}) S_e C(\underline{\theta})^*]$$

$$+ \mathrm{cov}(\langle C_j e_o, C(\underline{\theta})e_o\rangle, \langle C_\ell e_o, C(\underline{\theta})e_o\rangle)\};$$

$$(ii)\ E(z_1^{(q)} z_2^{(q)}) = -\sum_{j,\ell=1}^{m} r_j r_\ell \mathrm{tr}[S_e C(\underline{\theta})^* B A_j A(\underline{\theta}) \cdot (\sum_{\kappa=0}^{q-1} A(\underline{\theta})^\kappa S_v (A(\underline{\theta})^*)^\kappa) A_\ell^* B^*];$$

$$(iii)\ E(z_1^{(q)} z_k^{(q)}) = 0, \text{ falls } k \geq 3.$$

Beweis:

Zu (i): Definiere $\hat{z}_k^{(q)} := z_k^{(q)} - E(\langle C(\underline{\theta})e_o, D(\underline{r})e_o\rangle)$. Dann gilt:

$$E((\hat{z}_l^{(q)})^2) = E(\langle D(\underline{r})y_o^{(q)}, Bv_l + e_l\rangle^2)$$

$$-2E(\langle D(\underline{r})\tilde{y}_o^{(q)}, C(\underline{\theta})e_o\rangle\langle D(\underline{r})\tilde{y}_o^{(q)}, Bv_l + e_l\rangle) + E(\langle D(\underline{r})\tilde{y}_o^{(q)}, C(\underline{\theta})e_o\rangle^2)$$

$$= \mathrm{tr}[D(\underline{r})S_{\tilde{y}_o^{(q)}} D(\underline{r})^*(BS_vB^* + S_e)] + E([\langle D(\underline{r})\tilde{x}_o^{(q)}, C(\underline{\theta})e_o\rangle + \langle D(\underline{r})e_o, C(\underline{\theta})e_o\rangle]^2) =$$

$$\mathrm{tr}[D(\underline{r})(BS_{\tilde{x}_o^{(q)}}B^*+S_e)D(\underline{r})^*(BS_vB^*+S_e)]+\mathrm{tr}[D(\underline{r})BS_{\tilde{x}_o^{(q)}}B^*D(\underline{r})^*C(\underline{\theta})S_eC(\underline{\theta})^*]$$

$$+\ E(\langle D(\underline{r})e_o, C(\underline{\theta})e_o\rangle^2).$$

Der letzte Erwartungswert existiert aufgrund der Voraussetzung $E(\|e_o\|^4)$.

Wegen $D(\underline{r}) = \sum_{\ell=1}^{m} r_\ell C_\ell$, $S_{\tilde{x}_o^{(q)}} = \sum_{\ell=0}^{q} S_{A(\underline{\theta})^\ell v_{-\ell}} = \sum_{\ell=0}^{q} A(\underline{\theta})^\ell S_v (A(\underline{\theta})^*)^\ell$ und

$$E((z_1^{(q)})^2) = E((\hat{z}_1^{(q)})^2) - (E(\hat{z}_1^{(q)}))^2 = E((\hat{z}_1^{(q)})^2) - (E(\langle D(\underline{r})e_o, C(\underline{\theta})e_o\rangle))^2$$

folgt die Behauptung.

Zu (ii), (iii): Sei $k \geq 2$.

Gemäß Definition der Variablen $\hat{z}_\ell^{(q)}$ und der Beobachtungsgleichung (II) erhalten wir:

$$E(\hat{z}_1^{(q)}\hat{z}_k^{(q)}) =$$

$$E([\langle D(\underline{r})B\tilde{x}_o^{(q)}, Bv_1+e_1\rangle - \langle D(\underline{r})e_o, C(\underline{\theta})e_o\rangle + \langle D(\underline{r})e_o, Bv_1+e_1\rangle - \langle D(\underline{r})B\tilde{x}_o^{(q)}, C(\underline{\theta})e_o\rangle] \cdot$$

$$\cdot[\langle D(\underline{r})B\tilde{x}_{k-1}^{(q)}, Bv_k+e_k\rangle - \langle D(\underline{r})e_{k-1}, C(\underline{\theta})e_{k-1}\rangle + \langle D(\underline{r})e_{k-1}, Bv_k+e_k\rangle - \langle D(\underline{r})B\tilde{x}_{k-1}^{(q)}, C(\underline{\theta})e_{k-1}\rangle]).$$

Durch distributives Ausmultiplizieren erhält man eine Summe von Erwartungswerten, von denen aber viele verschwinden, da man einen unabhängigen, zentrierten Teil abspalten kann (vgl. dazu Rechenregel (b) in Lemma II.5.1); deshalb gilt

$$E(\hat{z}_1^{(q)}\hat{z}_k^{(q)}) = -E(\langle D(\underline{r})B\tilde{x}_o^{(q)}, Bv_1+e_1\rangle\langle D(\underline{r})B\tilde{x}_{k-1}^{(q)}, C(\underline{\theta})e_{k-1}\rangle)$$

$$+\ E(\langle D(\underline{r})e_o, C(\underline{\theta})e_o\rangle\ \langle D(\underline{r})e_{k-1}, C(\underline{\theta})e_{k-1}\rangle)$$

$$\underset{k\geq 2}{=} -E(\langle D(\underline{r})B\tilde{x}_o^{(q)}, Bv_1+e_1\rangle\langle D(\underline{r})B\tilde{x}_{k-1}^{(q)}, C(\underline{\theta})e_{k-1}\rangle)$$

$$+\ (E(\langle D(\underline{r})e_o, C(\underline{\theta})e_o\rangle))^2.$$

Falls $k \geq 3$ ist, fällt der erste Term wiederum gemäß Rechenregel (b) weg.

Für $k = 2$ nimmt dieser Term die Gestalt $E(\langle D(\underline{r})B\tilde{x}_o^{(q)}, e_1\rangle \langle D(\underline{r})B\tilde{x}_1^{(q)}, C(\underline{\theta})e_1\rangle)$ an.

Indem man die Definition von $\tilde{x}_o^{(q)}$ und $D(\underline{r})$ einsetzt und die Rechenregel (e), Lemma II.5.1, anwendet, ergibt sich schließlich $E(\hat{z}_1^{(q)} \cdot \hat{z}_2^{(q)}) =$

$$\sum_{j,\ell=1}^{m} r_j r_\ell \operatorname{tr}[S_e C(\underline{\theta})^* B A_j A(\underline{\theta}) \Big(\sum_{\kappa=0}^{q-1} (A(\underline{\theta}))^\kappa S_v (A(\underline{\theta})^*)^\kappa \Big) A_\ell^* B^*].$$

In beiden Fällten (ii), (iii) folgen die Resultate durch Übergang von den Variablen $\hat{z}_\ell^{(q)}$ zu $z_\ell^{(q)}$. □

Satz II.4.1 nimmt nun folgende Gestalt an.

Satz II.5.2:

Die Matrix $\Gamma := (\operatorname{tr}(B A_\ell F(\underline{\theta}) A_j^* B^*))_{\ell,j=1,\ldots,m}$ sei invertierbar; es gelte $E(\|v_o\|_H^2) < \infty$ und $E(\|e_o\|^4) < \infty$. Dann konvergiert $\sqrt{n}(\hat{\underline{\theta}}_n - \underline{\theta})$ schwach gegen die m-dimensionale Normalverteilung $\mathcal{N}(\underline{0}, \Gamma^{-1} \Phi (\Gamma^{-1})^*)$. Dabei besteht die Matrix Φ aus den Einträgen

$$\varphi_{\ell,j} =$$

$$\operatorname{tr}[(B A_j F(\underline{\theta}) A_\ell^* B^* + C_j S_e C_\ell^*)(B S_v B^* + S_e) + B A_j F(\underline{\theta}) A_\ell^* B^* C(\underline{\theta}) S_e C(\underline{\theta})^* - 2 S_e C(\underline{\theta})^* B A_j A(\underline{\theta}) F(\underline{\theta}) A_\ell^* B^*] +$$

$$+ \operatorname{cov}(\langle C_j e_o, C(\underline{\theta}) e_o \rangle, \langle C_\ell e_o, C(\underline{\theta}) e_o \rangle).$$

Beweis:

Gemäß Satz II.4.1 bleibt zu zeigen, daß die Bedingung $E(\|v_o\|_H^2) < \infty$ ausreicht (im Gegensatz zu $E(\|v_o\|_H^4) < \infty$) und zum andern, daß für die im Satz II.4.1 auftretenden Matrixelemente $\sigma_{\ell,j}^2 = \varphi_{\ell,j}$ gilt für alle $\ell,j=1,\ldots,m$.

Man weiß (vgl. den Beweis zu Lemma II.4.3), daß

(*) $\sigma_{\ell,j}^2 = \frac{1}{2} [\sigma^2(\underline{\theta}, \underline{e}_\ell + \underline{e}_j) - \sigma^2(\underline{\theta}, \underline{e}_\ell) - \sigma^2(\underline{\theta}, \underline{e}_j)]$ ist.

Andererseits gilt für alle $\underline{r} \in \mathbb{R}^m$ $\sigma^2(\underline{\theta}, \underline{r}) = \lim_{q \to \infty} \sigma^2(q, \underline{\theta}, \underline{r})$ und (vgl. Lemma II.4.1)

$$\sigma^2(q,\underline{\theta},\underline{r}) = E(z_1^{(q)}(\underline{r})^2) + 2 \cdot \sum_{k=2}^{\infty} E(z_1^{(q)}(\underline{r}) \cdot z_k^{(q)}(\underline{r}))$$

$$= E(z_1^{(q)}(\underline{r})^2) + 2 \cdot E(z_1^{(q)}(\underline{r}) \cdot z_2^{(q)}(\underline{r}))$$

$$= \sum_{j,\ell=1}^{m} r_j r_\ell \{ \mathrm{tr}[C_j (B(\sum_{\kappa=0}^{q} A(\underline{\theta})^\kappa S_v (A(\underline{\theta})^*)^\kappa) B^* + S_e) C_\ell^* (BS_v B^* + S_e)$$

$$+ C_j B(\sum_{\kappa=0}^{q} A(\underline{\theta})^\kappa S_v (A(\underline{\theta})^*)^\kappa) B^* C_\ell^* C(\underline{\theta}) S_e C(\underline{\theta})^*$$

$$- 2 S_e C(\underline{\theta})^* B A_j A(\underline{\theta}) \cdot (\sum_{\kappa=0}^{q-1} A(\underline{\theta})^\kappa S_v (A(\underline{\theta})^*)^\kappa) A_\ell^* B^*]$$

$$+ \mathrm{cov}(\langle C_j e_o, C(\underline{\theta}) e_o \rangle, \langle C_\ell e_o, C(\underline{\theta}) e_o \rangle) \}.$$

Somit erhalten wir:

$$\sigma^2(\underline{\theta},\underline{r}) = \sum_{j,\ell=1}^{m} r_j r_\ell \{ \mathrm{tr}[C_j (BF(\underline{\theta}) B^* + S_e) C_\ell^* (BS_v B^* + S_e)$$

$$+ C_j BF(\underline{\theta}) B^* C_\ell^* C(\underline{\theta}) S_e C(\underline{\theta})^* - 2 S_e C(\underline{\theta})^* B A_j A(\underline{\theta}) F(\underline{\theta}) A_\ell^* B^*]$$

$$+ \mathrm{cov}(\langle C_j e_o, C(\underline{\theta}) e_o \rangle, \langle C_\ell e_o, C(\underline{\theta}) e_o \rangle) \}$$

und deshalb wegen (*) und $C_j B = A_j B$, $j=1,\ldots,m$, die Behauptung $\sigma^2_{\ell,j} = \varphi_{\ell,j}$.—

Wir wenden uns nun der Abschwächung der Momentenbedingung zu. Dabei ist zu beachten, daß hier $\tilde{x}_o$ und $\tilde{x}_o - \tilde{x}_o^{(q-1)}$, $q \in \mathbb{N}$, stochastisch unabhängig von $Bv_1 + e_1 - C(\underline{\theta}) e_o$ sind. Dies hat zur Folge, daß die Abschätzungen am Beweisende von Lemma II.4.1 ohne die vorausgesetzte Existenz des vierten Moments von v_o durchgeführt werden können. Bei der Argumentation für die Endlichkeit von $E(z_1^2)$ benötigt man allerdings weiterhin $E(\|e_o\|^4) < \infty$. □

Zum Abschluß dieses Abschnitts schneiden wir noch zwei Problemkreise an, zum einen eine Abschwächung der Beobachtbarkeitsbedingung (O) (in II.1) und zum andern die Identifikation von zeitstetigen Prozessen, die durch eine stochastische Differentialgleichung gegeben sind und zeitdiskret beobachtet werden.

Falls die Voraussetzung (O) für das ursprüngliche dynamische Modell nicht erfüllt ist, kann man trotzdem in gewissen Fällen durch Übergang zum zugehörigen ℓ-Schritt-Problem die Systemparameter identifizieren.

Der Einfachheit halber behandeln wir nur den einparametrigen Fall und setzen auch $A_o = 0$. Wir nehmen also an, daß für ein $\mathbb{N} \ni \ell > 1$ A^ℓ (anstelle von A) die Beobachtbarkeitsbedingung (O) erfüllt. Dann betrachten wir $\hat{x}_k := x_{\ell \cdot k}$ anstelle von x_k für $k \in \mathbb{N}$. Offenbar gilt

$x_{k+\ell} = \theta^\ell A^\ell x_k + \sum_{j=1}^{\ell} (\theta A)^{\ell-j} v_{k+j}$ für alle $k \in \mathbb{N}$, sodaß man schreiben kann:

(+) $\hat{x}_{k+1} = \theta^\ell A^\ell \hat{x}_k + \hat{v}_{k+1}$, wobei $\hat{v}_{k+1} := \sum_{j=1}^{\ell} (\theta A)^{\ell-j} v_{k \cdot \ell + j}$.

Hier ist $(\hat{v}_k)_{k \in \mathbb{N}}$ eine Folge von unabhängigen, identisch verteilten Zufallsvariablen, da $(v_k)_{k \in \mathbb{N}}$ eine solche darstellt. Also liegt die Dynamik (+) genau in der Form der Systemgleichung (I) vor, nur mit der Ausnahme, daß die Systemstörungen $\hat{v}_k$ von dem unbekannten Parameter θ abhängen. Dadurch wird aber die Gültigkeit der Sätze II.5.1 und II.5.2 nicht beeinflußt, außer daß $S_{\hat{v}} = S_{\hat{v}}(\theta)$ nun von θ abhängt. Man beachte dazu auch, daß in der modifizierten Schätzformel (in der Situation der stochastischen Unabhängigkeit) keine Erwartungswerte bzgl. der v_k gebildet zu werden brauchen (vgl. den Anfang von § II.5). Außerdem gilt für das oben hergeleitete System wiederum die Stabilitätsbedingung, denn $\|\theta^\ell A^\ell\|_{\mathcal{L}(H)} \leq \|\theta A\|^\ell_{\mathcal{L}(H)} < 1$, da $\|\theta A\|_{\mathcal{L}(H)} < 1$. Also können die allgemein entwickelten Methoden angewendet werden, um θ^ℓ zu schätzen und damit auch θ selbst. Die Beobachtungsgleichung lautet hier natürlich

(++) $\hat{y}_k = \hat{x}_k + \hat{e}_k$ mit $\hat{y}_k := y_{\ell \cdot k}$ und $\hat{e}_k := e_{\ell \cdot k}$ für alle $k \in \mathbb{N}$. —

Abschließend führen wir ein Beispiel an für die Möglichkeit, die oben entwickelte Theorie auf die Identifikation zeitstetiger Prozesse, die diskret

gemessen werden, anzuwenden. Wir beschränken uns dabei auf die eindimensionale Situation. Die zeitliche Entwicklung des Systems sei gegeben durch

(I') $dx_t = \theta x_t dt + dw_t$, $t \in [0,\infty)$, $x_0 = 0$;

(II) $y_k = bx_{k\cdot\Delta} + e_k$, $k \in Z_+$.

Wie zuvor ist $(e_k)_{k\in Z_+}$ eine Folge von unabhängigen, identisch verteilten (reellen) Zufallsvariablen, $E(e_o) = 0$ und $E(e_o^2) < \infty$. $(w_t)_{t\in[0,\infty)}$ ist ein reellwertiger Standard-Wienerprozess, der unabhängig von $(e_k)_{k\in Z_+}$ ist. Δ ist eine beliebige positive Konstante, die sogenannte Abtastweite. Wir lösen das θ-Schätzproblem, indem wir für die Lösung $(x_t)_{t\in[0,\infty)}$ der stochastischen Differentialgleichung (I') eine (exakte) zeitdiskrete Rekursion angeben. Es gilt nämlich (s. [BR, '76])

$(\tilde{I})$ $x_{(k+1)\cdot\Delta} = \exp(\theta\Delta)\cdot x_{k\cdot\Delta} + v_{k+1}$, $k \in Z_+$, $x_0 = 0$,

wobei $v_k := \exp(k\cdot\Delta\theta)\cdot \int_{(k-1)\cdot\Delta}^{k\cdot\Delta} \exp(-s\theta)dw_s$ unabhängige, identisch verteilte, zentrierte gaußsche Variable sind. Wir können somit (falls $\theta < 0$) $\exp(\theta\Delta)$ schätzen und damit auch θ.

Kapitel III: Girsanovs Theorem im Hilbertraum und eine Anwendung auf ein statistisches Schätzproblem, welches mit einer Hilbertraum-wertigen stochastischen Differentialgleichung zusammenhängt

§ III.1: Girsanov-Sätze für Hilbertraum-wertige Wienerprozesse und stochastische Differentialgleichungen

In diesem Abschnitt führen wir unsere Verallgemeinerungen von Sätzen vom Girsanov-Typ an, und zwar für die Situation, in der der Zustandsraum der beteiligten stochastischen Prozesse ein separabler Hilbertraum ist. Zuerst studieren wir die Transformation eines unendlich-dimensionalen Wienerprozesses bei gleichzeitiger Veränderung des zugrundeliegenden Wahrscheinlichkeitsmaßes (Satz III.1.1). Anschließend werden wir mit Hilfe dieses Satzes ein Ergebnis über die Absolutstetigkeit von Maßen, die durch Lösungen von Hilbertraum-wertigen stochastischen Differentialgleichungen mit verschiedenen Drifttermen induziert werden, herleiten (Satz III.1.2). Letzteres Resultat wird im weiteren Verlauf des Kapitels III zur Konstruktion des Maximum-Likelihood-Schätzers für die parametrisierte Drift einer stochastischen Differentialgleichung herangezogen.

Im Gegensatz zur Situation im zweiten Kapitel setzen wir hier die direkte Beobachtungsmöglichkeit voraus. Demzufolge brauchen wir nur einen Hilbertraum H zu betrachten. Dieser sei reell und separabel. Mit $\langle \cdot, \cdot \rangle$ und $\| \cdot \|$ bezeichnen wir das zugehörige Skalarprodukt bzw. die Norm. $T > 0$ sei ein fixierter (End-) Zeitpunkt und $(w(t))_{t \in [0,T]}$ ein H-wertiger Wienerprozess. Darunter verstehen wir folgendes: $(w(t))_{t \in [0,T]}$ ist ein stochastischer Prozeß auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ mit Werten in H, der unabhängige und stationäre Zuwächse besitzt. Der Prozeß ist ferner zentriert und der zugehörige Kovarianz-

operator W ist nuklear, d.h. er besitzt endliche Spur. Außerdem sollen alle Trajektorien in 0 starten, d.h. $w(0) = 0$, und stetig sein. In Anlehnung an [CU/PR, '78] , S. 134, geben wir eine präzise

Definition III.1.1: [1]

Ein stochastischer Prozeß $w(t): \Omega \to H$, $t \in [0,T]$, heißt Wienerprozeß mit Kovarianzoperator W, wenn er folgende Bedingungen erfüllt:

(i) $E(\|w(t)\|^2) < \infty$ für alle $t \in [0,T]$;

(ii) $E(\langle w(t),x\rangle) = 0$ für alle $x \in H$, $t \in [0,T]$;

(iii) $E(\langle x,w(t)-w(s)\rangle \langle y,w(t)-w(s)\rangle) = (t-s)\langle Wx,y\rangle$
für alle $x,y \in H$, $s,t \in [0,T]$, $s \leq t$, wobei $W \in \mathcal{L}(H)$ ein nuklearer (nichtnegativ-definiter) Operator ist;

(iv) $w(t_4)-w(t_3)$ und $w(t_2)-w(t_1)$ sind stets stochastisch unabhängig, falls $0 \leq t_1 \leq t_2 \leq t_3 \leq t_4 \leq T$;

(v) die Abbildung $[0,T] \ni t \to w(t,\omega) \in H$ ist stetig für alle $\omega \in \Omega$;

(vi) $w(0) = 0$.

Im weiteren nehmen wir noch an, daß alle Eigenwerte λ_i, $i \in \mathbb{N}$, von W strikt positiv sind. Das ist insofern keine Einschränkung, als daß in den Eigenvektor-Koordinaten, die zu den Null-Eigenwerten gehören, keine Dynamik stattfindet. Diese Tatsache ersieht man aus folgender Darstellung des Wienerprozesses (gemäß Lemma 5.2.3, [CU/PR, '78]): $w(t) = \sum_{i=1}^{\infty} \beta_i(t) e_i$. Dabei ist $\{e_i : i \in \mathbb{N}\}$ ein vollständiges Orthonormalsystem, welches aus Eigenvektoren (bzgl. der Eigenwerte λ_i) des Kovarianzoperators W besteht. Die Prozesse $(\beta_i(t))_{t \in [0,T]} \equiv (\langle w(t),e_i\rangle)_{t \in [0,T]}$ sind gegenseitig unabhängige, reelle Wienerprozesse mit Varianzparameter λ_i, d.h. $E(\beta_i(t)^2) = \lambda_i t$, $t \in [0,T]$, $i \in \mathbb{N}$.

$\mathcal{F}_t = \sigma(w(s): 0 \leq s \leq t)$, $0 \leq t \leq T$, bezeichne diejenige Filtration, d.h. aufsteigende Folge von Sub-σ-Algebren in $(\Omega,\mathcal{A},P)$, die von dem Wienerprozeß

---

[1] Die Tatsache, daß $(w(t))_{t \in [0,T]}$ einen gaußschen Prozeß bildet, ist implizit in den Annahmen (i) bis (vi) enthalten.

$$E((Y_n(t_1,t_2)-Y_m(t_1,t_2))^2) = E((\sum_{i=m+1}^{n} \int_{t_1}^{t_2} \langle\varphi(s),e_i\rangle\, d(\frac{\beta_i(s)}{\sqrt{\lambda_i}}))^2)$$

$$= \sum_{i=m+1}^{n} E((\int_{t_1}^{t_2} \langle\varphi(s),e_i\rangle d(\frac{\beta_i(s)}{\sqrt{\lambda_i}}))^2) \leq \sum_{i=m+1}^{\infty} E(\int_0^T \langle\varphi(s),e_i\rangle^2 ds) \xrightarrow[m\to\infty]{} 0,$$

da $\sum_{i=1}^{\infty} E(\int_0^T \langle\varphi(s),e_i\rangle^2 ds) = E(\int_0^T \|\varphi(s)\|^2 ds) \leq \frac{C}{D}\cdot T < \infty.$

Schritt 3: Falls $t_2-t_1 \leq \frac{D}{18}$, ist $(\exp(2X_n(t_1,t_2)))_{n\in\mathbb{N}}$ gleichgradig integrierbar.—

Dies folgt sofort aus der Tatsache, daß $\sup_{n\in\mathbb{N}} E(\exp(3X_n(t_1,t_2))) < \infty$ ist.

Um letzteres zu beweisen, bemerken wir, daß man aufgrund der Hölderschen Ungleichung abschätzen kann:

$$E(\exp(3X_n(t_1,t_2)))$$

$$= E(\exp(3\cdot\int_{t_1}^{t_2} Q_n\varphi(s)dw_n(s) - \frac{27}{2}\int_{t_1}^{t_2} \|Q_n\varphi(s)\|^2 ds)\cdot\exp(12\int_{t_1}^{t_2} \|Q_n\varphi(s)\|^2 ds))$$

$$\leq [E(\exp(\int_{t_1}^{t_2} 9Q_n\varphi(s)dw_n(s) - \frac{1}{2}\int_{t_1}^{t_2} \|9Q_n\varphi(s)\|^2 ds))]^{\frac{1}{3}}\cdot[E(\exp(18\int_{t_1}^{t_2} \|Q_n\varphi(s)\|^2 ds))]^{\frac{2}{3}}.$$

Gemäß Theorem 1.1, [FR, '75], S. 152, ist der erste Faktor gleich 1, so daß es ausreicht zu zeigen, daß $E(\exp(18\int_{t_1}^{t_2} \|\varphi(s)\|^2 ds)) < \infty$ ist. Wegen der Jensenschen Ungleichung gilt

$$\exp(18\int_{t_1}^{t_2} \|\varphi(s)\|^2 ds) \leq \frac{1}{t_2-t_1}\int_{t_1}^{t_2} \exp(D\|\varphi(s)\|^2) ds.$$

Also folgt das Ergebnis in Anbetracht der Voraussetzung (1.1). -

Unmittelbar aus den beiden vorangehenden Schritten folgt

Schritt 4: Falls $t_2-t_1 \leq \frac{D}{18}$, liegt $\exp(X(t_1,t_2))$ im Raum $\mathcal{L}_2(\Omega,\mathcal{A},P)$ und $\mathcal{L}_2\text{-}\lim_{n\to\infty} \exp(X_n(t_1,t_2)) = \exp(X(t_1,t_2))$.—

Ergebnisse aus der Theorie der endlichdimensionalen Prozesse (vgl. [FR, '75], Lemma 1.3, S. 155 und Beweis des Theorems 1.1, S. 155) liefern

Schritt 5: Für $t_2-t_1 \leq \frac{D}{2}$ gilt $E(\exp(X_n(t_1,t_2))|\mathfrak{F}_{t_1}) = 1$ P-fast sicher.—

Das Bilden des bedingten Erwartungswertes ist gleichbedeutend mit einer Projektion im Raum $\mathcal{L}_2(\Omega,\mathcal{A},P)$. Also ist diese Operation stetig, so daß sich zusammen mit den Schritten 4 und 5 ergibt:

Schritt 6: Für $t_2-t_1 \leq \frac{D}{18}$ ist $E(\exp(X(t_1,t_2))|\mathfrak{F}_{t_1}) = 1$ P-fast sicher und insbesondere $E(\exp(X(t_1,t_2))) = 1$. -

Schritt 7: Für alle $0 \leq t_1 \leq t_2 \leq T$ liegt $\exp(X(t_1,t_2))$ im Raum $\mathcal{L}_1(\Omega,\mathcal{A},P)$ und es ist $E(\exp(X(t_1,t_2))|\mathfrak{F}_{t_1}) = 1$ P-fast sicher.
Insbesondere gilt $E(\exp X) = 1$.—

Angesichts der Schritte 1 und 2 liefert das Lemma von Fatou, daß $E(\exp(X(t_1,t_2))) \leq 1$ ist. Zerlegen wir nun das Intervall $[t_1,t_2]$ in Teilintervalle der Länge $\leq \frac{D}{18}$ (vgl. Schritt 6), erhalten wir mittels der gleichen Methode wie in [FR, '75], S. 155, 156, daß $E(\exp(X(t_1,t_2))|\mathfrak{F}_{t_1}) = 1$ ist P-fast sicher.—

Die in Lemma III.1.1 behauptete $\mathcal{L}_1$-Konvergenz folgt nun aus dem Schefféschen Lemma. Überdies gilt:

$$|\tilde{P}_{n'}(A)-\tilde{P}(A)| \leq \int_A |\exp X_{n'} - \exp X|dP$$

$$\leq \|\exp X_{n'} - \exp X\|_1 \xrightarrow[n\to\infty]{} 0. \quad \square$$

Wir können nun das erste Hauptergebnis dieses Abschnitts formulieren.

Satz III.1.1: Sei H ein reeller separabler Hilbertraum mit Norm $\|\cdot\|$ und $(w(t))_{t\in[0,T]}$ ein H-wertiger Wienerprozeß. Der zugehörige nukleare Kovarianzoperator W habe nur strikt positive Eigenwerte λ_i, $i\in\mathbb{N}$, und $(e_i)_{i\in\mathbb{N}}$ sei ein vollständiges Orthonormalsystem in H, das aus den Eigenvektoren von W besteht.

Man betrachte einen $(w(t))_{t\in[0,T]}$-adaptierten Prozeß $(\varphi(t))_{t\in[0,T]}$, der folgende Eigenschaft besitzt:

(1.1) Es existieren strikt positive Konstanten C und D derart, daß für alle $t\in[0,T]$

$$E(\exp(D\|\varphi(t)\|^2)) \leq C < \infty.$$

Ferner definiere man für alle $t\in[0,T]$

$$(1.2)\quad \tilde{w}(t) = w(t) - \int_0^t \sqrt{W}\,\varphi(s)ds.$$

Dann ist $(\tilde{w}(t))_{t\in[0,T]}$ ein H-wertiger Wienerprozeß auf $(\Omega,\mathcal{A},\tilde{P})$ mit Kovarianzoperator W. Dabei ist das Wahrscheinlichkeitsmaß $\tilde{P}$ festgelegt durch

$$\frac{d\tilde{P}}{dP} = \exp(\mathcal{L}_2\text{-}\lim_{n\to\infty} \int_0^T Q_n\varphi(s)d(\Lambda_n Q_n w(t)) - \frac{1}{2}\int_0^T \|\varphi(s)\|^2 ds).$$

Hier bezeichnet Q_n die Projektion von H auf den abgeschlossenen Teilraum, der durch $e_1,\ldots,e_n$ erzeugt wird. Die Matrix Λ_n ist gegeben durch

$$(\Lambda_n)_{i,j} = \frac{1}{\sqrt{\lambda_i}}\,\delta_{i,j},\quad i,j=1,\ldots,n.$$

Beweis: Offensichtlich besitzt $\tilde{w}$ stetige Trajektorien, und es gilt $\tilde{w}(0) = 0$. Wir beweisen, daß $\tilde{w}$ bzgl. des Maßes $\tilde{P}$ die gleiche Verteilung auf $(H^{[0,T]}, \bigotimes_{t\in[0,T]} \mathcal{B}(H))$ induziert wie der Prozeß w bzgl. P. Hierbei bezeichnet $\mathcal{B}(H)$ die Borel-σ-Algebra in H. Wir zeigen, daß für alle Zeitpunkte $0 \leq t_1 < t_2 < \ldots < t_k \leq T$ gilt:

$$\tilde{P}\circ(\tilde{w}(t_1),\ldots,\tilde{w}(t_k)) = P\circ(w(t_1),\ldots,w(t_k)).$$

Angesichts der Tatsache, daß $\lim_{n\to\infty} \|Q_n x - x\| = 0$ ist für alle $x\in H$, genügt es zu verifizieren, daß für alle $n\in\mathbb{N}$ und alle Borel-Mengen $A_1,\ldots,A_k$ des $\mathbb{R}^k$

$$\tilde{P}(Q_n\tilde{w}(t_1)\in A_1,\ldots,Q_n\tilde{w}(t_k)\in A_k) = P(Q_n w(t_1)\in A_1,\ldots,Q_n w(t_k)\in A_k).$$

Indem wir Lemma III.1.1 und das endlichdimensionale Girsanov-Theorem (vgl. Th. 7.2.1, [FR, '75]) anwenden, können wir für die linke Seite in obiger Gleichung schreiben (sei $m' \geq n$ stets):

$$\lim_{m\to\infty} \tilde{P}_m, (Q_n\tilde{w}(t_1)\in A_1,\ldots,Q_n\tilde{w}(t_k)\in A_k)$$

$$= \lim_{m\to\infty} \tilde{P}_m, (\Lambda_m, Q_m, \tilde{w}(t_1)\in \Lambda_m, (\mathbb{R}^{m'-n}\times A_1),\ldots,\Lambda_m, Q_m, \tilde{w}(t_k)\in \Lambda_m, (\mathbb{R}^{m'-n}\times A_k))$$

$$= \lim_{m\to\infty} P_m, (\Lambda_m, Q_m, w(t_1)\in \Lambda_m, (\mathbb{R}^{m'-n}\times A_1),\ldots,\Lambda_m, Q_m, w(t_k)\in \Lambda_m, (\mathbb{R}^{m'-n}\times A_k))$$

$$= \lim_{m\to\infty} P_m, (Q_n w(t_1)\in A_1,\ldots,Q_n w(t_k)\in A_k).$$

Der letzte Ausdruck ist aber identisch mit der rechten Seite, womit der Beweis von Satz III.1.1 abgeschlossen ist. □

Bemerkung: Unser Ergebnis in Satz III.1.1 entspricht den Theoremen 3.5 und 3.6 in [OU, '73] (Ouvrards Voraussetzungen sind insofern etwas stärker als unsere Bedingung (1.1), denn er verlangt $D^2 > 3 + 2\sqrt{2}$). Beim Vergleich der beiden Resultate sollte man beachten, daß unser Prozeß φ zu dem Prozeß φ_o, mit dem Ouvrard arbeitet, in der Beziehung $\varphi = \sqrt{W}\,\varphi_o$ steht.—

Wir möchten nun den Satz III.1.1 heranziehen, um Ergebnisse über die Absolutstetigkeit von gewissen Maßen im Raum $(C([0,T],H),\mathcal{B})$ zu erzielen. Diese Maße werden induziert durch Lösungen von H-wertigen stochastischen Differentialgleichungen (SDGL) mit verschiedenen Drifttermen. Das entsprechende Vorgehen in der Situation eines endlichdimensionalen Zustandsraumes ist u.a. auf den Seiten 166 bis 169 in [FR, '75] dargestellt. Mit $C([0,T],H)$ bezeichnen wir den Banachraum aller stetigen Funktionen $f: [0,T] \to H$, der mit der Supremumsnorm ausgestattet ist. $\mathcal{B}$ ist die zugehörige Borel- -Algebra. Unendlichdimensionale SDGL werden interpretiert im Sinne von [CU/PR, '78], S. 143 ff; d.h. Lösungen von SDGL sind Lösungen der entsprechenden Integralgleichungen. H-wertige stochastische Integrale werden in zwei Schritten

konstruiert. Zunächst wird der das Differential bildende Wienerprozeß in seine $(e_i)_{i\in\mathbb{N}}$-Koordinaten entwickelt. Anschließend wird das stochastische Integral eines H-wertigen Integranden und eines reellen stochastischen Differentials wie üblich durch Approximation mittels Treppenfunktionen gebildet (Einzelheiten sind in [CU/PR, '78], Def. 5.25 und Lemma 5.26, S. 136 dargestellt; eine allgemeinere Theorie H-wertiger stochastischer Integrale ist z.B. in [KUN, '70] formuliert).

Im folgenden arbeiten wir nur mit solchen Evolutionsgleichungen, die eine eindeutige starke Lösung besitzen (vgl. [CU/PR, '78], Def. 5.30). Dafür hinreichende Bedingungen an die in der SDGL auftretenden Operatoren findet man z.B. im Theorem 5.35 [CU/PR, '78]. Für $t > 0$ führen wir den Raum ein

$$\mathcal{B}^2(0,t;\mathcal{L}(H)) =$$

$$\{\phi:[0,t]\to\mathcal{L}(H)\,|\,\phi \text{ ist meßbar bzgl. } (\mathcal{L}(H),\|\cdot\|_{\mathcal{L}(H)}) \text{ und } \int_0^t \|\phi(s)\|^2_{\mathcal{L}(H)}\,ds < \infty\},$$

wobei $(\mathcal{L}(H), \|\cdot\|_{\mathcal{L}(H)})$ den Banachraum der beschränkten linearen Operatoren in H bezeichnet. Wir geben nun das zweite Hauptergebnis dieses Abschnitts an.

Satz III.1.2: Zusätzlich zur in Satz III.1.1 verwendeten Notation seien A und $\tilde{A}$ infinitesimale Erzeugende stark stetiger Halbgruppen in H derart, daß der Definitionsbereich von A in demjenigen des Operators $\tilde{A}$ enthalten ist. Seien $D \in \mathcal{B}^2(0,T;\mathcal{L}(H))$ und $z(t)$, $\tilde{z}(t)$ Lösungen von

(1.3) $dz(t) = Az(t)dt + D(t)dw(t)$, $0 \le t \le T$, $z(0) = z_0 \in H$,

bzw.

(1.4) $d\tilde{z}(t) = \tilde{A}\tilde{z}(t)dt + D(t)dw(t)$, $0 \le t \le T$, $\tilde{z}(0) = z_0$.

Es existiere ein Prozeß $\varphi: [0,T] \times \Omega \to H$, der die Bedingung (1.1) erfüllt und auch

(1.5) $D(t)\sqrt{W}\,\varphi(t) = \tilde{A}z(t) - Az(t)$ für alle $t\in[0,T]$ P-fast sicher.

Dann induzieren $z = (z(t))_{t\in[0,T]}$ und $\tilde{z} = (\tilde{z}(t))_{t\in[0,T]}$ äquivalente Maße $\mu = P\bullet z$ und $\tilde{\mu} = P\bullet\tilde{z}$ in $(C([0,T],H),\mathfrak{B})$ und die Radon-Nikodym-Ableitung ist gegeben durch

$$\frac{d\tilde{\mu}}{d\mu}((z(t,\omega))_{t\in[0,T]}) = \exp(\mathfrak{L}_2\text{-}\lim_{n\to\infty}\int_0^T Q_n\varphi(t,\omega)d(\Lambda_n Q_n w(t,\omega)) - \frac{1}{2}\int_0^T \|\varphi(t,\omega)\|^2 dt).$$

Beweis: Für $\tilde{w}(s) := w(s) - \int_0^s \sqrt{W}\,\varphi(u)du$, $0\le s\le t\le T$ erhalten wir gemäß Voraussetzung (1.5)

$$\int_0^t D(s)d\tilde{w}(s) = \int_0^t D(s)dw(s) + \int_0^t Az(s)ds - \int_0^t \tilde{A}z(s)ds.$$

Also löst z(t) die stochastische Differentialgleichung

$$dz(t) = \tilde{A}z(t)dt + D(t)d\tilde{w}(t),\ z(0) = z_0.$$

In Anbetracht des Satzes III.1.1 sind z(t) und $\tilde{z}(t)$ Lösungen der gleichen stochastischen Differentialgleichung, aber mit verschiedenen zugrundeliegenden Wienerprozessen, d.h. Wahrscheinlichkeitsmaßen P bzw. $\tilde{P}$, also $P\bullet\tilde{z} = \tilde{P}\bullet z$. Aufgrund einer allgemeingültigen Transformationsregel für Dichten beim Übergang zum Bildmaß, nämlich $\frac{d\tilde{P}\bullet z}{dP\bullet z}(z(\omega)) = \frac{d\tilde{P}}{dP}(\omega)$ für P-fast alle $\omega\in\Omega$, folgt die Behauptung sofort mittels der expliziten Formel für $\frac{d\tilde{P}}{dP}$ aus Satz III.1.1. □

Korollar III.1.1: Sei A infinitesimale Erzeugende einer stark stetigen Halbgruppe in H und z(t) sei Lösung der Gleichung

$$dz(t) = Az(t)dt + dw(t),\ 0\le t\le T,\ z(0) = z_0\in H.$$

Es mögen zwei strikt positive Konstanten C, D existieren derart, daß

(1.6) $E(\exp(D(\sum_{i=1}^{\infty}\frac{1}{\lambda_i}\langle Az(t),e_i\rangle^2))) \le C$ für alle $t\in[0,T]$ gilt.

Dann induzieren $(z(t))_{t\in[0,T]}$ und $(w(t)+z_0)_{t\in[0,T]}$ äquivalente Maße ν und μ_{z_0} auf $(C([0,T],H),\mathfrak{B})$ und

$$\frac{d\nu}{d\mu_{z_0}}((z(t,\omega))_{t\in[0,T]}) =$$

$$\exp\Big(\frac{1}{2}\sum_{i=1}^{\infty}\frac{1}{\lambda_i}\int_0^T \langle Az(t,\omega),e_i\rangle^2 dt + \mathfrak{L}_2\text{-}\lim_{n\to\infty}\sum_{i=1}^{n}\frac{1}{\lambda_i}\int_0^T \langle Az(t,\omega),e_i\rangle d(\langle w(t,\omega),e_i\rangle)=$$

$$\exp\Big(-\frac{1}{2}\sum_{i=1}^{\infty}\frac{1}{\lambda_i}\int_0^T \langle Az(t,\omega),e_i\rangle^2 dt + \mathfrak{L}_2\text{-}\lim_{n\to\infty}\sum_{i=1}^{n}\frac{1}{\lambda_i}\int_0^T \langle Az(t,\omega),e_i\rangle d(\langle z(t,\omega),e_i\rangle)\Big).$$

Beweis: Es handelt sich hier um den Spezialfall $D(t) = id_H$ für alle $t\in[0,T]$ und $A = 0$. Darüberhinaus ist Voraussetzung (1.6) äquivalent mit (1.1) und (1.5), und es gilt $\frac{d\nu}{d\mu_{z_0}} = \left(\frac{d\mu_{z_0}}{d\nu}\right)^{-1}$. □

Dann induzieren $z = (z(t))_{t\in[0,T]}$ und $\tilde{z} = (\tilde{z}(t))_{t\in[0,T]}$ äquivalente Maße $\mu = P\bullet z$ und $\tilde{\mu} = P\bullet\tilde{z}$ in $(C([0,T],H),\mathfrak{B})$ und die Radon-Nikodym-Ableitung ist gegeben durch

$$\frac{d\tilde{\mu}}{d\mu}((z(t,\omega))_{t\in[0,T]}) = \exp(\mathfrak{L}_2\text{-}\lim_{n\to\infty}\int_0^T Q_n\varphi(t,\omega)d(\Lambda_n Q_n w(t,\omega)) - \frac{1}{2}\int_0^T \|\varphi(t,\omega)\|^2 dt).$$

Beweis: Für $\tilde{w}(s) := w(s) - \int_0^s \sqrt{W}\,\varphi(u)du$, $0\le s\le t\le T$ erhalten wir gemäß Voraussetzung (1.5)

$$\int_0^t D(s)d\tilde{w}(s) = \int_0^t D(s)dw(s) + \int_0^t Az(s)ds - \int_0^t \tilde{A}z(s)ds.$$

Also löst z(t) die stochastische Differentialgleichung

$$dz(t) = \tilde{A}z(t)dt + D(t)d\tilde{w}(t),\ z(0) = z_o.$$

In Anbetracht des Satzes III.1.1 sind z(t) und $\tilde{z}(t)$ Lösungen der gleichen stochastischen Differentialgleichung, aber mit verschiedenen zugrundeliegenden Wienerprozessen, d.h. Wahrscheinlichkeitsmaßen P bzw. $\tilde{P}$, also $P\bullet\tilde{z} = \tilde{P}\bullet z$. Aufgrund einer allgemeingültigen Transformationsregel für Dichten beim Übergang zum Bildmaß, nämlich $\frac{d\tilde{P}\bullet z}{dP\bullet z}(z(\omega)) = \frac{d\tilde{P}}{dP}(\omega)$ für P-fast alle $\omega\in\Omega$, folgt die Behauptung sofort mittels der expliziten Formel für $\frac{d\tilde{P}}{dP}$ aus Satz III.1.1. □

Korollar III.1.1: Sei A infinitesimale Erzeugende einer stark stetigen Halbgruppe in H und z(t) sei Lösung der Gleichung

$$dz(t) = Az(t)dt + dw(t),\ 0\le t\le T,\ z(0) = z_o\in H.$$

Es mögen zwei strikt positive Konstanten C, D existieren derart, daß

(1.6) $E(\exp(D(\sum_{i=1}^{\infty}\frac{1}{\lambda_i}\langle Az(t),e_i\rangle^2))) \le C$ für alle $t\in[0,T]$ gilt.

Dann induzieren $(z(t))_{t\in[0,T]}$ und $(w(t)+z_o)_{t\in[0,T]}$ äquivalente Maße ν und μ_{z_o} auf $(C([0,T],H),\mathfrak{B})$ und

$$\frac{d\nu}{d\mu_{z_o}}((z(t,\omega))_{t\in[0,T]}) =$$

$$\exp\Big(\frac{1}{2}\sum_{i=1}^{\infty}\frac{1}{\lambda_i}\int_0^T \langle Az(t,\omega),e_i\rangle^2 dt + \mathfrak{L}_2\text{-}\lim_{n\to\infty}\sum_{i=1}^{n}\frac{1}{\lambda_i}\int_0^T \langle Az(t,\omega),e_i\rangle d(\langle w(t,\omega),e_i\rangle)=$$

$$\exp\Big(-\frac{1}{2}\sum_{i=1}^{\infty}\frac{1}{\lambda_i}\int_0^T \langle Az(t,\omega),e_i\rangle^2 dt + \mathfrak{L}_2\text{-}\lim_{n\to\infty}\sum_{i=1}^{n}\frac{1}{\lambda_i}\int_0^T \langle Az(t,\omega),e_i\rangle d(\langle z(t,\omega),e_i\rangle)\Big).$$

Beweis: Es handelt sich hier um den Spezialfall $D(t) = id_H$ für alle $t\in[0,T]$ und $A = 0$. Darüberhinaus ist Voraussetzung (1.6) äquivalent mit (1.1) und (1.5), und es gilt $\frac{d\nu}{d\mu_{z_o}} = \left(\frac{d\mu_{z_o}}{d\nu}\right)^{-1}$. □

§ III.2: Parameterschätzung für den Driftterm einer unendlichdimensionalen stochastischen Differentialgleichung

Wir behandeln hier das statistische Problem, den Driftterm einer H-wertigen stochastischen Differentialgleichung zu identifizieren. Genauer gesagt untersuchen wir Gleichungen des Typs

(2.1) $dz(t) = A(\theta)z(t)dt + dw(t)$, $z(0) = z_o \in H$.

Dabei gehört θ einer Parametermenge Θ an, einem Intervall in $\mathbb{R}$. Für alle $\theta \in \Theta$ ist $A(\theta)$ infinitesimale Erzeugende einer stark stetigen Halbgruppe in H derart, daß (2.1) eine eindeutige starke Lösung besitzt. Das Problem besteht darin, den (unbekannten) Parameter θ mittels einer beobachteten Trajektorie des Prozesses z(t) zu schätzen. Um explizite Berechnungen zu ermöglichen, nehmen wir an, daß $A(\theta) = \theta A$ gilt mit einem bekannten Operator A. Wir verlangen zunächst, daß die Bedingung (1.6) aus Korollar III.1.1 erfüllt ist. Dann gilt:

$$\frac{d\nu_\theta}{d\mu_{z_o}}((z(t,\omega))_{t\in[0,T]}) =$$

$$\exp\Big(-\frac{1}{2}\theta^2 \sum_{i=1}^{\infty} \frac{1}{\lambda_i} \int_0^T \langle Az(t,\omega), e_i\rangle^2 dt + \theta \cdot \mathcal{L}_2\text{-}\lim_{n\to\infty} \Big(\sum_{i=1}^{n} \frac{1}{\lambda_i} \int_0^T \langle Az(t,\omega), e_i\rangle d(\langle z(t,\omega), e_i\rangle)\Big)$$

wobei ν_θ das von $(z(t))_{t\in[0,T]} \equiv (z_\theta(t))_{t\in[0,T]}$, der Lösung der Gleichung (2.1), induzierte Maß auf $(C([0,T],H), \mathfrak{B})$ ist und μ_{z_o} dem Prozeß $(w(t)+z_o)_{t\in[0,T]}$ entspricht.

Der Maximum-Likelihood-Schätzer $\hat{\theta}_T$ ist also gegeben durch

$$(2.2)\quad \hat{\theta}_T(\omega) = \frac{\mathcal{L}_2\text{-}\lim_{n\to\infty} \sum_{i=1}^{n} \frac{1}{\lambda_i} \int_0^T \langle Az(t,\omega), e_i\rangle d(\langle z(t,\omega), e_i\rangle)}{\sum_{i=1}^{\infty} \frac{1}{\lambda_i} \int_0^T \langle Az(t,\omega), e_i\rangle^2 dt},$$

falls der Nenner größer als Null ist.

Im folgenden untersuchen wir asymptotische Eigenschaften des Schätzers $\hat{\theta}_T$ für $T \to \infty$. Zuerst weisen wir die starke Konsistenz des Schätzers nach, d.h. $\hat{\theta}_T$ konvergiert P-fast sicher gegen den wahren Parameter θ mit $T \to \infty$. Wir bemerken noch, daß wir nun die Bedingung (1.6) aus Korollar III.1.1 nicht mehr benötigen. Sie diente lediglich dazu, die Schätzformel $\hat{\theta}_T$, (2.2), durch die Maximum-Likelihood-Methode zu motivieren. Bevor wir die Konsistenz beweisen, formulieren wir ein Lemma, das eine Verallgemeinerung des Theorems 1.7 [BA/RA, '80], S. 402, darstellt.

Lemma III.2.1:

Sei $(G_i(t))_{t\in[0,\infty)}$, $i\in\mathbb{N}$, eine Familie (bzgl. $(w(t))_{t\in[0,\infty)}$) nichtvorgreifender reellwertiger stochastischer Prozesse auf $(\Omega,\mathcal{A},P)$ mit den Eigenschaften:

(2.3) Es existiert ein $n_o\in\mathbb{N}$ mit $\sum_{i=1}^{n_o} \int_0^\infty G_i(t)^2 dt = \infty$ P-fast sicher

und

(2.4) $E(\sum_{i=1}^{\infty} \int_0^L G_i(t)^2 dt) < \infty$ für alle $L\in(0,\infty)$.

Für $T\in[0,\infty)$ definiere man die Stopzeiten τ_T durch

$$\tau_T(\omega) := \begin{cases} \inf\{u\in[0,\infty): \sum_{i=1}^{\infty} \int_0^u G_i(t,\omega)^2 dt > T\} \\ +\infty, \text{ falls } \{\ldots\} = \phi \end{cases}$$

Dann konvergiert für alle nichtnegativen T

$$f_n(T) := \sum_{i=1}^{n} \frac{1}{\sqrt{\lambda_i}} \int_0^{\tau_T} G_i(t)\, d(\langle w(t), e_i\rangle) \text{ in } \mathcal{L}_2(\Omega,\mathcal{A},P) \text{ mit } n \to \infty.$$

Der Grenzprozeß $(f(T))_{T\in[0,\infty)}$ ist ein Standard-Wienerprozeß (d. h. ein Wienerprozeß mit der Diffusionskonstanten 1).

Beweis: Wir definieren für alle $n\in\mathbb{N}$, $T\in[0,\infty)$ Stopzeiten
=======

$$\tau_{T,n}(\omega) := \begin{cases} \inf\{u\in[0,\infty): \sum_{i=1}^{n} \int_0^u G_i(t,\omega)^2 dt > T\} \\ +\infty, \text{ falls } \{\ldots\} = \emptyset \end{cases}.$$

Angesichts der Voraussetzung (2.3) sind τ_T und $\tau_{T,n}$ P-fast sicher endlich für alle $n \geq n_o$, $T\in[0,\infty)$. Offensichtlich konvergiert $\tau_{T,n}$ monoton fallend gegen τ_T. Wir kürzen ab: $\eta_i(t) := \frac{1}{\sqrt{\lambda_i}} < w(t), e_i >$, $i\in\mathbb{N}$; d.h. $\eta_i(t)$ sind gegenseitig unabhängige Standard-Wienerprozesse. Alsdann setzen wir für $T\in[0,\infty)$, $n \geq n_o$

$$k_n(T) = \sum_{i=1}^{n} \int_0^{\tau_{T,n}} G_i(t)d\eta_i(t).$$

Gemäß Theorem 1.7 [BA/RA, '80], S. 402, sind die Prozesse $(k_n(T))_{T\in[0,\infty)}$ für alle $n\in\mathbb{N}$ Standard-Wienerprozesse. Außerdem konvergiert $f_n(T) - k_n(T)$ gegen 0 im $\mathcal{L}_2$-Sinn mit $n \to \infty$, denn

$$E((k_n(T)-f_n(T))^2) = \sum_{i=1}^{n} E([\int_T^{\tau_{T,n}} G_i(t)d\eta_i(t)]^2) = E(\int_{\tau_T}^{\tau_{T,n}} (\sum_{i=1}^{n} G_i(t)^2)dt).$$

Die Behauptung folgt aus dem Satz von der majorisierten Konvergenz, weil $\int_{\tau_T}^{\tau_{T,n}} (\sum_{i=1}^{n} G_i(t)^2)dt \leq T$ ist.

Wir zeigen nun, daß $(f_n(T))_{n\in\mathbb{N}}$ eine Cauchy-Folge in $\mathcal{L}_2(\Omega,\mathcal{A},P)$ ist, deren Limes wir mit $f(T)$ bezeichnen. Für $n > m$ erhalten wir nämlich

$$E((f_n(T)-f_m(T))^2) = \sum_{i=m+1}^{n} E(\int_0^{\tau_T} G_i(t)^2 dt)$$

$$\leq \sum_{i=m+1}^{\infty} E(\int_0^{\tau_T} G_i(t)^2 dt) \xrightarrow[m\to\infty]{} 0, \text{ da}$$

$$\sum_{i=1}^{\infty} E(\int_0^{\tau_T} G_i(t)^2 dt) = E(T) = T < \infty.$$

Zusammenfassend können wir feststellen, daß für alle $T\in[0,\infty)$

$\mathcal{L}_2\text{-}\lim_{n\to\infty} k_n(T) = f(T)$ ist. Insbesondere konvergieren die endlichdimensionalen Randverteilungen der Prozesse $(k_n(T))_{T\in[0,\infty)}$ schwach gegen diejenigen von

$(f(T))_{T\in[0,\infty)}$. Aber die endlichdimensionalen Verteilungen von $(k_n(T))_{T\in[0,\infty)}$ hängen gar nicht von n ab. Somit folgt das Ergebnis. □

Satz III.2.1 (Starke Konsistenz des Schätzers $\hat{\theta}_T$):

Sei $z(t) = z_\theta(t)$ die Lösung der folgenden stochastischen Differentialgleichung:

$$dz(t) = \theta Az(t)dt + dw(t),\ z(0) = z_o,\ t\in[0,\infty).$$

Wir setzen voraus, daß

$$(2.5)\quad \sum_{i=1}^{n_o} \frac{1}{\lambda_i} \int_0^\infty \langle Az(t),e_i\rangle^2 dt = \infty \text{ ist P-fast sicher für ein } n_o\in\mathbb{N}$$

und daß

$$(2.6)\quad E\Big(\sum_{i=1}^{\infty} \frac{1}{\lambda_i} \int_0^L \langle Az(t),e_i\rangle^2 dt\Big) < \infty \text{ ist für alle } L\in(0,\infty).$$

Dann ist der Schätzer

$$\hat{\theta}_T(\omega) = \frac{\mathcal{L}_2\text{-}\lim\limits_{n\to\infty} \sum\limits_{i=1}^{n} \frac{1}{\lambda_i} \int_0^T \langle Az(t,\omega),e_i\rangle d(\langle z(t,\omega),e_i\rangle)}{\sum\limits_{i=1}^{\infty} \frac{1}{\lambda_i} \int_0^T \langle Az(t,\omega),e_i\rangle^2 dt}$$

stark konsistent, d.h. $\lim\limits_{T\to\infty} \hat{\theta}_T = \theta$ P-fast sicher.

Beweis: Zunächst bemerken wir, daß in Anbetracht von (2.5) der Schätzer $\hat{\theta}_T$ zumindest für hinreichend große Zeitpunkte T P-fast sicher wohldefiniert ist. Wegen $d(\langle z(t),e_i\rangle) = \theta \cdot \langle Az(t),e_i\rangle dt + d(\langle w(t),e_i\rangle)$ können wir schreiben:

$$\hat{\theta}_T = \theta + \frac{\mathcal{L}_2\text{-}\lim\limits_{n\to\infty} \sum\limits_{i=1}^{n} \frac{1}{\lambda_i} \int_0^T \langle Az(t),e_i\rangle d(\langle w(t),e_i\rangle)}{\sum\limits_{i=1}^{\infty} \frac{1}{\lambda_i} \int_0^T \langle Az(t),e_i\rangle^2 dt}$$

$$=: \theta + g(T).$$

Indem wir die Stopzeiten

$$\tau_T(\omega) := \begin{cases} \inf \{u\in\mathbb{R}_+ : \sum_{i=1}^{\infty} \frac{1}{\lambda_i} \int_0^u \langle Az(t,\omega), e_i\rangle^2 dt > T\} \\ + \infty \text{ , falls } \{\ldots\} = \emptyset \end{cases} \quad , \; T\in[0,\infty),$$

einführen, erhalten wir durch Anwendung von Lemma III.2.1

$(G_i(t) \equiv \frac{1}{\sqrt{\lambda_i}} \langle Az(t), e_i\rangle)$ und des starken Gesetzes der großen Zahlen für den Wienerprozeß, daß $\lim_{T\to\infty} g(\tau_T(\omega),\omega) = 0$ ist für P-fast alle ω in Ω. Gemäß der Voraussetzung (2.6) konvergiert τ_T monoton gegen $+\infty$ P-fast sicher mit $T \to \infty$. Überdies werden durch $(g(T,\omega))_{T\in[0,\infty)}$ und $(g(\tau_T(\omega),\omega))_{T\in[0,\infty)}$ die gleichen Pfade beschrieben für P-fast alle $\omega\in\Omega$, die allerdings mit verschiedenen Geschwindigkeiten durchlaufen werden. Deshalb gilt auch $\lim_{T\to\infty} g(T) = 0$ P-fast sicher und damit die Behauptung. □

Um die Bedeutung der Voraussetzungen (2.5) und (2.6) zu erläutern, geben wir zwei Beispielklassen von Operatoren an, die diese Bedingungen erfüllen.

Beispiel III.2.1: Die orthonormierten Vektoren e_i, $i\in\mathbb{N}$, seien gleichzeitig Eigenvektoren des Operators A mit zugehörigen Eigenwerten β_i, $i\in\mathbb{N}$. Falls A ein Hilbert-Schmidt-Operator ist und $\sup_{i\in\mathbb{N}} \frac{1}{\lambda_i} E(\langle z_o, e_i\rangle^2) < \infty$, dann gilt (2.6). Im Fall $A \neq 0$ ist (2.5) erfüllt.

Beweis: Der Einfachheit halber nehmen wir an, daß $z_o = 0$ und $\beta_i \neq 0$ ist für alle $i\in\mathbb{N}$. Wegen $\langle Ax, e_i\rangle = \langle x, Ae_i\rangle = \beta_i\langle x, e_i\rangle$ für alle $x\in H$, $i\in\mathbb{N}$ erhalten wir

$$E\Big(\sum_{i=1}^{\infty} \frac{1}{\lambda_i} \int_0^L \langle Az(t), e_i\rangle^2 dt\Big) = \sum_{i=1}^{\infty} \frac{\beta_i^2}{\lambda_i} \int_0^L E(\langle z(t), e_i\rangle^2) dt.$$

Überdies ist es leicht einzusehen, daß

$\langle z(t), e_i\rangle = \int_0^t \exp(\beta_i\theta(t-s)) d(\langle w(s), e_i\rangle)$ ist, woraus folgt daß

$$E(\langle z(t), e_i\rangle^2) = \frac{\lambda_i}{2\beta_i\theta} (\exp(2\beta_i\theta t) - 1).$$

Deshalb gilt

$$E\Big(\sum_{i=1}^{\infty} \frac{1}{\lambda_i} \int_0^L \langle Az(t), e_i\rangle^2 dt\Big) = \frac{1}{4\theta^2} \sum_{i=1}^{\infty} (\exp(2\beta_i \theta L) - 2\beta_i \theta L - 1).$$

Der letzte Term ist aber genau dann endlich, wenn $\sum_{i=1}^{\infty} \beta_i^2 < \infty$, was gerade bedeutet, daß A ein Hilbert-Schmidt-Operator ist. Ebenso ist es leicht einzusehen, daß die Voraussetzung (2.5) genau dann erfüllt ist, wenn mindestens einer der Eigenwerte β_i nicht verschwindet (vgl. dazu die Gleichungen (17.44) und (17.57) in [LI/SH, '78]).

Beispiel III.2.2: Wir konstruieren hier einen beschränkten Operator A, der die Voraussetzungen (2.5) und (2.6) erfüllt, aber kein Hilbert-Schmidt-Operator ist. Dazu wählen wir eine Teilfolge $k: \mathbb{N} \to \mathbb{N}$, $i < k(i) < k(i+1)$ für alle $i \in \mathbb{N}$ derart, daß $\lambda := \sum_{i=1}^{\infty} \frac{\lambda_{k(i)}}{\lambda_i} < \infty$ ist und definieren $Ax = \sum_{i=1}^{\infty} \langle x, e_{k(i)}\rangle e_i$ für alle $x \in H$. Mittels einer einfachen Rechnung erhalten wir (vgl. Lemma III.2.3):

$$E(\langle Az(t), e_i\rangle^2) = \frac{1}{\theta} \sum_{j=1}^{\infty} \lambda_j \int_0^{\theta \cdot t} \langle \exp(sA) Ae_j, e_i\rangle^2 ds$$

$$= \frac{1}{\theta} \sum_{n=0}^{\infty} \sum_{\ell=0}^{\infty} \sum_{j=1}^{\infty} \lambda_j \int_0^{\theta \cdot t} \frac{s^{n+\ell}}{n!\ell!} \delta_{j,k^{n+1}(i)} \cdot \delta_{j,k^{\ell+1}(i)} ds$$

$$= \frac{1}{\theta} \sum_{n=0}^{\infty} \lambda_{k^{n+1}(i)} \cdot \int_0^{\theta \cdot t} \frac{s^{2n}}{(n!)^2} ds \leq \frac{\lambda_{k(i)}}{\theta} \cdot \int_0^{\theta \cdot t} \exp(s^2) ds$$

$$\leq \lambda_{k(i)} \cdot t \cdot \exp(\theta^2 t^2).$$

Dabei bezeichnet für $j \in \mathbb{N}$ k^j die j-fache Iterierte der Abbildung $k: \mathbb{N} \to \mathbb{N}$.

Deshalb gilt:

$$E\Big(\sum_{i=1}^{\infty} \frac{1}{\lambda_i} \int_0^L \langle Az(t), e_i\rangle^2 dt\Big) \leq \lambda \cdot \int_0^L t \exp(\theta^2 t^2) dt < \infty, \text{ d.h. (2.6).}$$

Um (2.5) zu verifizieren, beweisen wir, daß

(+) $\lim_{T \to \infty} E(\exp(-\int_0^T \langle Az(t), e_i\rangle^2 dt)) = 0$ ist für alle $i \in \mathbb{N}$.

Durch Anwendung der Jensenschen Ungleichung bzgl. der Gleichverteilung auf dem Intervall [0,T] erhält man

$$E(\exp(-\int_0^T <Az(t),e_i>^2 dt)) \leq E(\frac{1}{T}\int_0^T \exp(-T<Az(t),e_i>^2)dt)$$

$$\leq \frac{1}{T}\int_0^T E(\exp(-t<Az(t),e_i>^2)dt.$$

Ferner gilt für eine beliebige zentrierte gaußsche Zufallsvariable η die Beziehung $E(\exp(-\eta^2)) = (2E(\eta^2)+1)^{-1/2}$. Also kann man die rechte Seite in der obigen Ungleichung ersetzen durch $\frac{1}{T}\int_0^T (2tE(<Az(t),e_i>^2)+1)^{-1/2}\,dt$. Aufgrund der Abschätzung $E(<Az(t),e_i>^2) \geq \lambda_{k(i)}\cdot t$ (vgl. die Herleitung von (2.6)) folgt (+).—

Im weiteren Verlauf des Abschnitts leiten wir (unter gewissen Zusatzvoraussetzungen) die asymptotische Normalität des Schätzers $\hat{\theta}_T$ her. Der Beweis des folgenden Lemmas (ein Zentraler Grenzwertsatz für stochastische Integrale) basiert auf den Ideen von Yu.A. Kutoyants (s. [KUT, '75] und auch [BA/RA, '80], S. 402).

Lemma III.2.2:

Die Familie $(G_i(t))_{t\in[0,\infty)}$, $i\in\mathbb{N}$, nicht-vorgreifender Prozesse erfülle die Bedingungen (2.3) und (2.4) aus Lemma III.2.1. Überdies konvergiere $\frac{1}{T}\sum_{i=1}^{\infty}\int_0^T G_i(t)^2dt$ gegen eine positive Konstante c P-stochastisch mit $T\to\infty$. Dann konvergiert $\frac{1}{\sqrt{T}}\ \mathfrak{L}_2\text{-}\lim_{n\to\infty}\sum_{i=1}^{n}\frac{1}{\sqrt{\lambda_i}}\int_0^T G_i(t)d(<w(t),e_i>)$ schwach gegen die Normalverteilung $\mathcal{N}(0,c)$ mit Mittelwert 0 und Varianz c mit $T\to\infty$.

Beweis: Für $T>0$ sei $T^* = T(1+2c)$, und man definiere für alle $i\in\mathbb{N}$ Prozesse

$$G_i^*(t) = \begin{cases} G_i(t), & \text{falls } t\in[0,T]\cup(T^*,\infty), \\ 1, & \text{falls } t\in(T,T^*] \end{cases}$$

und Stopzeiten

$$\tau_{T,c}(\omega) = \inf\{u\in\mathbb{R}_+: \sum_{i=1}^{\infty}\int_0^u G_i^*(t,\omega)^2dt > c\cdot T\}$$

$$= \inf\{u\in\mathbb{R}_+: \sum_{i=1}^{\infty}\int_0^u (\frac{1}{\sqrt{c}}\ G_i^*(t,\omega))^2 > T\}.$$

Man beachte, daß $\int_0^{T^*} G_i^*(t)^2 dt \geq \int_T^{T^*} 1 dt > c \cdot T$ ist, also $\tau_{T,c} \leq T^*$ für alle $T > 0$.

Gemäß Lemma III.2.1 ist

$$f_c(T) := \mathcal{L}_2\text{-}\lim_{n\to\infty} \sum_{i=1}^{n} \frac{1}{\sqrt{\lambda_i}} \int_0^{\tau_{T,c}} \frac{1}{\sqrt{c}} G_i^*(t) d(\langle w(t), e_i \rangle)$$

$\mathcal{N}(0,T)$-verteilt.

Deshalb ist $g_c(T) := \frac{1}{\sqrt{T}} \mathcal{L}_2\text{-}\lim_{n\to\infty} \sum_{i=1}^{n} \frac{1}{\sqrt{\lambda_i}} \int_0^{\tau_{T,c}} G_i^*(t) d(\langle w(t), e_i \rangle)$

$\mathcal{N}(0,c)$-verteilt.

Man setze für $T > 0$ $h(T) = \frac{1}{\sqrt{T}} \mathcal{L}_2\text{-}\lim_{n\to\infty} \sum_{i=1}^{n} \frac{1}{\sqrt{\lambda_i}} \int_0^T G_i(t) d(\langle w(t), e_i \rangle)$.

Es bleibt somit zu zeigen, daß $h(T) - g_c(T)$ P-stochastisch gegen Null konvergiert mit $T \to \infty$. Für beliebige positive Zahlen δ und α können wir abschätzen

$$P(|h(T) - g_c(T)| > \delta)$$

$$= P\Big(\frac{1}{\sqrt{T}} \Big| \mathcal{L}_2\text{-}\lim_{n\to\infty} \Big(\sum_{i=1}^{n} \frac{1}{\sqrt{\lambda_i}} \Big(\int_0^T G_i(t) d(\langle w(t), e_i \rangle) - \int_0^{\tau_{T,c}} G_i^*(t) d(\langle w(t), e_i \rangle) \Big) \Big) \Big| > \delta \Big)$$

$$= P\Big(\frac{1}{\sqrt{T}} \Big| \mathcal{L}_2\text{-}\lim_{n\to\infty} \sum_{i=1}^{n} \frac{1}{\sqrt{\lambda_i}} \int_0^{T^*} (1_{[0,T]}(t) - 1_{[0,\tau_{T,c}]}(t)) G_i^*(t) d(\langle w(t), e_i \rangle) \Big| > \delta \Big)$$

$$\leq \liminf_{n\to\infty} P\Big(\frac{1}{\sqrt{T}} \Big| \sum_{i=1}^{n} \frac{1}{\sqrt{\lambda_i}} \int_0^{T^*} (1_{[0,T]}(t) - 1_{[0,\tau_{T,c}]}(t)) G_i^*(t) d(\langle w(t), e_i \rangle) \Big| > \frac{\delta}{2} \Big)$$

$$\leq \frac{4\alpha}{\delta^2} + \liminf_{n\to\infty} P\Big(\frac{1}{T} \sum_{i=1}^{n} \frac{1}{\lambda_i} \int_0^{T^*} |1_{[0,T]}(t) - 1_{[0,\tau_{T,c}]}(t)| G_i^*(t)^2 dt \geq \alpha \Big)$$

$$\leq \frac{4\alpha}{\delta^2} + P\Big(\frac{1}{T} \sum_{i=1}^{\infty} \frac{1}{\lambda_i} \int_0^{T^*} |1_{[0,T]}(t) - 1_{[0,\tau_{T,c}]}(t)| G_i^*(t)^2 dt \geq \alpha \Big),$$

wobei wir eine n-dimensionale Version des Theorems 2.6 [FR, '75], S. 63, angewendet haben. Wir müssen deshalb beweisen, daß die Wahrscheinlichkeit des Ereignisses

$$A(T) := \Big\{ \frac{1}{T} \sum_{i=1}^{\infty} \frac{1}{\lambda_i} \int_0^{T^*} |1_{[0,T]}(t) - 1_{[0,\tau_{T,c}]}(t)| G_i^*(t)^2 dt \geq \alpha \Big\}$$

gegen Null konvergiert mit $T \to \infty$. Angesichts der Definition von G_i^* und $\tau_{T,c}$ erhalten wir

$$A(T)\cap\{\tau_{T,c}\leq T\}=\{\frac{1}{T}\sum_{i=1}^{\infty}\frac{1}{\lambda_i}\int_0^T G_i(t)^2dt - c\geq\alpha\}\cap\{\tau_{T,c}\leq T\}$$

$$\subseteq \{|\frac{1}{T}\sum_{i=1}^{\infty}\frac{1}{\lambda_i}\int_0^T G_i(t)^2dt-c|\geq\alpha\}\cap\{\tau_{T,c}\leq T\}$$

und ähnlich

$$A(T)\cap\{\tau_{T,c}>T\}\subseteq\{|c-\frac{1}{T}\sum_{i=1}^{\infty}\frac{1}{\lambda_i}\int_0^T G_i(t)^2dt|\geq\alpha\}\cap\{\tau_{T,c}>T\},$$

zusammenfassend also

$$A(T)\subseteq\{|\frac{1}{T}\sum_{i=1}^{\infty}\frac{1}{\lambda_i}\int_0^T G_i(t)^2dt - c|\geq\alpha\}\quad.$$

Aufgrund der Voraussetzung des Lemmas folgt $\lim_{T\to\infty} P(A(T)) = 0$. □

Satz III.2.2: (Asymptotische Normalität des Maximum-Likelihood-Schätzers $\hat{\theta}_T$):

$z(t) = z_\theta(t)$ erfülle die Voraussetzungen (2.5) und (2.6) des Satzes III.2.1.

Ferner gelte

$$(2.7)\quad \lim_{T\to\infty}\frac{1}{T}\sum_{i=1}^{\infty}\frac{1}{\lambda_i}\int_0^T \langle Az(t),e_i\rangle^2dt = c \quad P\text{-stochastisch},$$

wobei c eine positive Konstante ist.

Dann ist der Schätzer $\hat{\theta}_T$ asymptotisch normal; es gilt nämlich, daß $\sqrt{T}\,(\hat{\theta}_T-\theta)$ schwach gegen die Normalverteilung $\mathcal{N}(0,\frac{1}{c})$ konvergiert mit $T\to\infty$.

Beweis:

Es ist (vgl. den Beweis des Satzes III.2.1)

$$\sqrt{T}(\hat{\theta}_T-\theta) = \frac{\frac{1}{\sqrt{T}}\,\mathfrak{L}_2\text{-}\lim_{n\to\infty}\sum_{i=1}^{n}\frac{1}{\lambda_i}\int_0^T\langle Az(t),e_i\rangle d(\langle w(t),e_i\rangle)}{\frac{1}{T}\sum_{i=1}^{\infty}\frac{1}{\lambda_i}\int_0^T\langle Az(t),e_j\rangle^2dt}\quad.$$

Gemäß Voraussetzung konvergiert der Nenner P-stochastisch gegen c, und wegen Lemma III.2.2 konvergiert der Zähler in Verteilung gegen $\mathcal{N}(0,c)$. Demgemäß folgt

die Behauptung aufgrund eines Standardresultats aus der Theorie der schwachen Konvergenz (vgl. [GÄ/ST, '77], Kor. 8.66). □

Zum Ende dieses Abschnitts diskutieren wir die Voraussetzung (2.7) des Satzes III.2.2 und geben dafür hinreichende Bedingungen an. Offensichtlich ist (2.7) eine Forderung an das ergodische Grenzverhalten des Prozesses $\sum_{i=1}^{\infty} \frac{1}{\lambda_i} \langle Az(t), e_i \rangle^2$.

Zunächst charakterisieren wir die Verteilung von $z(t)$, $t \in [0,\infty)$, der (starken) Lösung der stochastischen Differentialgleichung $dz(t) = Az(t)dt + dw(t), z(0) = z_0$, wobei z_0 eine gaußsche H-wertige Zufallsgröße ist, die unabhängig ist von $\sigma(w(t): t \in [0,\infty))$, der σ-Algebra, die von dem in der Differentialgleichung auftretenden Wienerprozeß erzeugt wird. Dabei heißt eine Zufallsvariable $\eta: (\Omega, \mathcal{A}, P) \to (H, \mathcal{B}(H))$ gaußsch, falls $\langle \eta, x \rangle$ für alle $x \in H$ normalverteilt ist (vgl. [KU, '75], § 2). Mit $(T_t)_{t \in [0,\infty)}$ bezeichnen wir die stark stetige Operatorenhalbgruppe, die von A erzeugt wird. $E(\eta)$ für eine Hilbertraum- oder Banachraumwertige Zufallsvariable η auf $(\Omega, \mathcal{A}, P)$ ist der Erwartungswert bzgl. des Maßes P(,wobei das Integral natürlich im Bochnerschen Sinn zu verstehen ist). Überdies bezeichne für $y, z \in H$ $y \bullet z \in \mathcal{L}(H)$ denjenigen Operator, für den $y \bullet z(h) = y \langle z, h \rangle$ gilt für alle $h \in H$.

Lemma III.2.3:

Für alle $t \in [0,\infty)$ ist $z(t)$ ein gaußsches Zufallselement. Zusätzlich gelten die Formeln:

(2.8) $E(z(t)) = T_t E(z_0)$

und

(2.9) $E(z(t) \bullet z(t)) = E((T_t z_0) \bullet (T_t z_0)) + \int_0^t T_u W T_u^* du.$

Beweis:

Es gilt $z(t) = T_t z_o + \int_0^t T_{t-s} dw(s)$, wobei die Summanden stochastisch unabhängig sind (vgl. [CU/PR, '78], § 5.3). $z(t)$ ist also gaußsch, da sowohl $T_t z_o$ als auch $\int_0^t T_{t-s}\, dw(s)$ gaußsche Elemente sind. Dabei folgt die Tatsache, daß ein stochastisches Integral mit einem (deterministischen linearen) Integranden stets gaußsch ist, unmittelbar aus der Konstruktion eines H-wertigen stochastischen Integrals (vgl. z.B. [CU/PR, '78], S. 136,137). Die Formeln (2.8) und (2.9) resultieren direkt aus Lemma 5.37, [CU/PR, '78]. □

Von nun an arbeiten wir nur noch mit exponentiell stabilen Halbgruppen $(T_t)_{t\in[0,\infty)}$, d.h. wir nehmen an, daß Konstanten $M \geq 1$ und $\lambda > 0$ existieren derart, daß

(2.10) $\|T_t\|_{\mathcal{L}(H)} \leq M \exp(-\lambda t)$ für alle $t \in [0,\infty)$ gilt.

Kriterien an den infinitesimalen Erzeuger A, die die Stabilität der zugehörigen Halbgruppe $(T_t)_{t\geq 0}$ sicherstellen, hat z.B. Slemrod aufgestellt, [SL, '76].

Mit $(\mathcal{L}_1(H), \|\cdot\|_1)$ bezeichnen wir den Banachraum aller nuklearen Operatoren, d.h. aller Operatoren $S \in \mathcal{L}(H)$ mit endlicher Spur, der mit der sogenannten Spurklassennorm (engl. trace class norm) ausgestattet ist (vgl. [KU, '75], S. 9ff.).

Man definiere

$\mathcal{S} := \{S \in \mathcal{L}_1(H)$: S ist nichtnegativ-definit und selbstadjungiert$\}$,

und für $x \in H$, $S \in \mathcal{S}$ sei $\mathcal{N}(x,S)$ das gaußsche Maß auf $(H, \mathcal{B}(H))$ mit Mittelwert x und Kovarianzoperator S. Es gilt nun folgendes

Lemma III.2.4:

a) $S_\infty := \int_0^\infty T_u W T_u^* du$ existiert in $\mathcal{L}_1(H)$

und $\int_0^t T_u W T_u^*$ konvergiert gegen S_∞ in $(\mathcal{L}_1(H), \|\cdot\|_1)$ mit $t \to \infty$.

b) Sei $\tilde{z}_o$ ein $\mathcal{N}(0,S_\infty)$-verteiltes Zufallselement in H, das unabhängig von $\sigma(w(s): s\in[0,\infty))$ ist.

Dann ist der Prozeß $\tilde{z}(t) = T_t\tilde{z}_o + \int_0^t T_{t-s}dw(s)$ ein strikt stationärer ergodischer Markov-Prozeß.

Beweis:

Durch Anwendung des Theorems 1.6 [KU, '75] erhalten wir

$$\left\| \int_t^\infty T_u W T_u^* du \right\|_1 \le \int_t^\infty \|T_u W T_u^*\|_1 du \le \int_t^\infty \|T_u\|_{\mathcal{L}(H)} \cdot \|W\|_1 \|T_u^*\|_{\mathcal{L}(H)} du$$

$$\le M^2 \cdot \|W\|_1 \int_t^\infty e^{-2\lambda u} du = (2\lambda)^{-1} M^2 \|W\|_1 e^{-2\lambda t} .$$

Also folgt a). —

Für $t\in[0,\infty)$, z_o = const. $\in$ H und Borelmengen $B \in \mathcal{B}(H)$ definiere man die Übergangswahrscheinlichkeiten

$$P_t(z_o,B) = \mathcal{N}(T_t z_o, (T_t z_o \bullet T_t z_o) + \int_0^t T_u W T_u^* du)(B).$$

Gemäß Lemma III.2.3 beschreibt $P_t(z_o,\cdot)$ die Verteilung der Variablen $z(t,z_o) = T_t z_o + \int_0^t T_{t-s}dw(s)$, d.h. der Lösung der stochastischen Differentialgleichung $dz(t) = Az(t)dt + dw(t)$, $z(0) = z_o$. Insbesondere ist $(P_t)_{t\in[0,\infty)}$ eine Markovsche Halbgruppe (vgl. [AR/CU/KO, '80], Th. 4.1). In Anbetracht von Lemma III.2.3 kann leicht nachgeprüft werden, daß $\tilde{z}(t)$ $\mathcal{N}(0,S_\infty)$-verteilt ist für alle Zeitpunkte $t\in[0,\infty)$. Wegen

$$\|z(t,z_o)-\tilde{z}(t)\| = \|T_t(z_o-\tilde{z}_o)\| \le M\|z_o-\tilde{z}_o\| e^{-\lambda t}$$

erhalten wir für alle $\varepsilon > 0$ $\lim_{t\to\infty} P(\|z(t,z_o)-\tilde{z}(t)\| > \varepsilon) = 0$, d.h. $z(t,z_o)$ und $\tilde{z}(t)$ sind stochastisch äquivalent.

Letzteres impliziert, daß für alle $z_o\in H$ die Übergangswahrscheinlichkeiten

$P_t(z_o, \cdot)$ schwach gegen $\mathcal{N}(0,S_\infty)$ konvergieren mit $t \to \infty$. Dies bedeutet, daß $\mathcal{N}(0,S_\infty)$ die einzige invariante Verteilung bzg. der Markovschen Halbgruppe $(P_t)_{t \in [0,\infty)}$ ist. Mittels einer leichten Abänderung des Beweises für den Fall zeitdiskreter Prozesse (vgl. z.B. [BI, '61], S. 52) kann man die Ergodizität des Prozesses $(\tilde{z}(t))_{t \in [0,\infty)}$ herleiten. □

Eine einfache Rechnung ergibt

Lemma III.2.5:

Seien f und g nichtnegative Borel-meßbare (reellwertige) Funktionen auf $[0,\infty)$ derart, daß

$$\text{(i)} \quad \limsup_{T\to\infty} \frac{1}{T} \int_0^T f(t)dt < \infty$$

und (ii) g beschränkt und $\lim_{T\to\infty} g(T) = 0$ ist.

Dann gilt:

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T g(t)f(t)dt = 0. \; -$$

Für $\theta > 0$ sei $z(t,z_o,\theta)$ die Lösung der stochastischen Differentialgleichung $dz(t) = \theta Az(t)dt + dw(t)$, $z(0) = z_o = \text{const.} \in H$, $t \in [0,\infty)$, wobei A die infinitesimale Erzeugende der stark stetigen (exponentiell stabilen) Halbgruppe $(T_t)_{t \in [0,\infty)}$ ist. Mit S_∞ bezeichnen wir $\int_0^\infty T_u W T_u^* du$ und $S_\infty(\theta) = \frac{1}{\theta} S_\infty$. Sei ferner D(A) der Definitionsbereich des Operators A. Im folgenden Lemma sind nun Kriterien dafür angegeben, daß Lösungen $z(t,z_o,\theta)$ die Voraussetzung (2.7) des Satzes über die asymptotische Normalität des Schätzers $\hat{\theta}_T$ erfüllen.

Lemma III.2.6:

Alle Vektoren e_i, $i \in \mathbb{N}$, seien Eigenvektoren des Operators A, und es gelte

$$\sum_{i=1}^{\infty} \frac{1}{\lambda_i} \langle Az_o, e_i \rangle^2 < \infty.$$

1) Falls $\mathcal{N}(0,S_\infty(\theta))(D(A)) = 1$ und $c_1(\theta) := \sum_{i=1}^{\infty} \frac{1}{\lambda_i} \int_H \langle Ax,e_i\rangle^2 \mathcal{N}(0,S_\infty(\theta))(dx) < \infty$

ist, konvergiert $\frac{1}{T} \sum_{i=1}^{\infty} \frac{1}{\lambda_i} \int_0^T \langle Az(t,z_o,\theta),e_i\rangle^2 dt$ P-fast sicher gegen $c_1(\theta)$

mit $T \to \infty$.

2) Falls $e_i \in D(A^*)$ für alle $i \in \mathbb{N}$ und $c_2(\theta) := \sum_{i=1}^{\infty} \frac{1}{\lambda_i} \int_H \langle x,A^*e_i\rangle^2 \mathcal{N}(0,S_\infty(\theta))(dx) < \infty$

ist, konvergiert $\frac{1}{T} \sum_{i=1}^{\infty} \frac{1}{\lambda_i} \int_0^T \langle Az(t,z_o,\theta),e_i\rangle^2 dt$ P-fast sicher gegen $c_2(\theta)$

mit $T \to \infty$.

Beweis:

Wir bemerken zunächst, daß $D(A) = \{x \in H: \lim_{h \downarrow 0} \frac{1}{h} (T_h x - x) \text{ existiert}\} \in \mathcal{B}(H)$ ist.

Man definiere $\tilde{z}(t,\theta) = T_{\theta t}\tilde{z}_o(\theta) + \int_0^t T_{\theta(t-s)} dw(s)$, wobei die Startvariable

$\tilde{z}_o(\theta)$ $\mathcal{N}(0,S_\infty(\theta))$-verteilt und stochastisch unabhängig von $\sigma(w(s): s \in [0,\infty))$

ist. Offenbar gilt $S_\infty(\theta) = \int_0^\infty T_{\theta u} W T^*_{\theta u} du$, und gemäß Lemma III.2.4 ist

$(\tilde{z}(t,\theta))_{t \in [0,\infty)}$ ein ergodischer Prozeß. Also ist auch für jede Borel-meßbare

Funktion $\varphi: H \to \mathbb{R}$ $(\varphi(\tilde{z}(t,\theta)))_{t \in [0,\infty)}$ ein ergodischer Prozeß, sodaß im Fall

$E(|\varphi(\tilde{z}(0,\theta))|) < \infty$ der (zeitkontinuierliche) Ergodensatz Gültigkeit besitzt:

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T \varphi(\tilde{z}(t,\theta))dt = E(\varphi(\tilde{z}(0,\theta))) \text{ P-fast sicher.}$$

Mit der Wahl $\varphi_1(x) := \begin{cases} \sum_{i=1}^{\infty} \frac{1}{\lambda_i} \langle Ax,e_i\rangle^2 & , \text{falls } x \in D(A) \\ 0 & , \text{falls } x \in H \setminus D(A) \end{cases}$

bzw. $\varphi_2(x) := \sum_{i=1}^{\infty} \frac{1}{\lambda_i} \langle x,A^*e_i\rangle^2$, $x \in H$, erhalten wir im Fall 1 bzw. 2

$$\text{(2.11)} \quad \lim_{T\to\infty} \frac{1}{T} \sum_{i=1}^{\infty} \frac{1}{\lambda_i} \int_0^T \langle A\tilde{z}(t,\theta),e_i\rangle^2 dt = c_1(\theta) \text{ P-fast sicher}$$

bzw.

$$\text{(2.12)} \quad \lim_{T\to\infty} \frac{1}{T} \sum_{i=1}^{\infty} \frac{1}{\lambda_i} \int_0^T \langle \tilde{z}(t,\theta),A^*e_i\rangle^2 dt = c_2(\theta) \text{ P-fast sicher.}$$

Aufgrund der Hölderschen Ungleichung kann man abschätzen:

$$\left|\frac{1}{T}\sum_{i=1}^{\infty}\frac{1}{\lambda_i}\int_0^T \langle Az(t,z_o,\theta),e_i\rangle^2 dt - \frac{1}{T}\sum_{i=1}^{\infty}\frac{1}{\lambda_i}\int_0^T \langle A\tilde{z}(t,\theta),e_i\rangle^2 dt\right|$$

$$= \left|\frac{1}{T}\int_0^T\left(\sum_{i=1}^{\infty}\frac{1}{\lambda_i}\langle A(z(t,z_o,\theta)-\tilde{z}(t,\theta)),e_i\rangle\langle A(z(t,z_o,\theta)+\tilde{z}(t,\theta)),e_i\rangle\right)dt\right|$$

$$\leq \frac{1}{T}\int_0^T\left(\sum_{i=1}^{\infty}\frac{1}{\lambda_i}\langle A(z(t,z_o,\theta)-\tilde{z}(t,\theta)),e_i\rangle^2\right)^{1/2}\cdot\left(\sum_{i=1}^{\infty}\frac{1}{\lambda_i}\langle A(z(t,z_o,\theta)+\tilde{z}(t,\theta)),e_i\rangle^2\right)^{\frac{1}{2}}dt$$

bzw. (im Fall 2)

$$\left|\frac{1}{T}\sum_{i=1}^{\infty}\frac{1}{\lambda_i}\int_0^T \langle Az(t,z_o,\theta),e_i\rangle^2 dt - \frac{1}{T}\sum_{i=1}^{\infty}\frac{1}{\lambda_i}\int_0^T \langle \tilde{z}(t,\theta),A^*e_i\rangle^2 dt\right|$$

$$= \left|\frac{1}{T}\int_0^T\left(\sum_{i=1}^{\infty}\frac{1}{\lambda_i}\langle z(t,z_o,\theta)-\tilde{z}(t,\theta),A^*e_i\rangle\langle z(t,z_o,\theta)+\tilde{z}(t,\theta),A^*e_i\rangle\right)dt\right|$$

$$\leq \frac{1}{T}\int_0^T\left(\sum_{i=1}^{\infty}\frac{1}{\lambda_i}\langle z(t,z_o,\theta)-\tilde{z}(t,\theta),A^*e_i\rangle^2\right)^{1/2}\cdot\left(\sum_{i=1}^{\infty}\frac{1}{\lambda_i}\langle z(t,z_o,\theta)+\tilde{z}(t,\theta),A^*e_i\rangle^2\right)^{\frac{1}{2}}dt.$$

Angesichts der Beziehungen (2.11), (2.12) und des Lemmas III.2.5 bleibt zu zeigen:

(2.13) $\lim\limits_{t\to\infty}\sum\limits_{i=1}^{\infty}\frac{1}{\lambda_i}\langle A(z(t,z_o,\theta)-\tilde{z}(t,\theta)),e_i\rangle^2 = 0$ P-fast sicher

bzw.

(2.14) $\lim\limits_{t\to\infty}\sum\limits_{i=1}^{\infty}\frac{1}{\lambda_i}\langle z(t,z_o,\theta)-\tilde{z}(t,\theta),A^*e_i\rangle^2 = 0$ P-fast sicher

und

(2.15) $\limsup\limits_{T\to\infty}\frac{1}{T}\int_0^T\left(\sum\limits_{i=1}^{\infty}\frac{1}{\lambda_i}\langle A(z(t,z_o,\theta)+\tilde{z}(t,\theta)),e_i\rangle^2\right)^{1/2}dt < \infty$ P-fast sicher

bzw.

(2.16) $\limsup\limits_{T\to\infty}\frac{1}{T}\int_0^T\left(\sum\limits_{i=1}^{\infty}\frac{1}{\lambda_i}\langle z(t,z_o,\theta)+\tilde{z}(t,\theta),A^*e_i\rangle^2\right)^{1/2}dt < \infty$ P-fast sicher.

Es ist $z(t,z_o,\theta) = T_{\theta t}z_o + \int_0^t T_{\theta(t-s)}dw(s)$, und da $T_{\theta t}$ und A auf D(A) kommutieren,

lassen sich die Terme in (2.13) bzw. (2.14) umformen zu

$$\sum_{i=1}^{\infty} \frac{1}{\lambda_i} \langle T_{\theta t} A(z_o - \tilde{z}_o(\theta)), e_i \rangle^2 \text{ bzw. } \sum_{i=1}^{\infty} \frac{1}{\lambda_i} \langle T_{\theta t}(z_o - \tilde{z}_o(\theta)), A^* e_i \rangle^2.$$

Weil die Vektoren e_i, $i \in \mathbb{N}$, nach Voraussetzung Eigenvektoren von A sind und somit auch bzgl. $T_{\theta t}$, erhalten wir für alle $x \in H$:

$$\sum_{i=1}^{\infty} \frac{1}{\lambda_i} \langle T_{\theta t} x, e_i \rangle^2 \leq \| T_{\theta t} \|^2_{\mathcal{L}(H)} \cdot \sum_{i=1}^{\infty} \frac{1}{\lambda_i} \langle x, e_i \rangle^2$$

$$\text{bzw. } \sum_{i=1}^{\infty} \frac{1}{\lambda_i} \langle T_{\theta t} x, A^* e_i \rangle^2 \leq \| T_{\theta t} \|^2_{\mathcal{L}(H)} \cdot \sum_{i=1}^{\infty} \frac{1}{\lambda_i} \langle x, A^* e_i \rangle^2.$$

In Anbetracht der einfachen Ungleichung

$$(2.17)\quad \langle x+y, e_i \rangle^2 \leq 2(\langle x, e_i \rangle^2 + \langle y, e_i \rangle^2) \text{ für } x, y \in H,$$

gilt ferner

$$\sum_{i=1}^{\infty} \frac{1}{\lambda_i} \langle A(z_o - \tilde{z}_o(\theta)), e_i \rangle^2 \leq 2 \cdot \sum_{i=1}^{\infty} \frac{1}{\lambda_i} \langle A z_o, e_i \rangle^2 + 2 \cdot \sum_{i=1}^{\infty} \frac{1}{\lambda_i} \langle A \tilde{z}_o(\theta), e_i \rangle^2$$

$$\text{bzw. } \sum_{i=1}^{\infty} \frac{1}{\lambda_i} \langle z_o - \tilde{z}_o(\theta), A^* e_i \rangle^2 \leq 2 \cdot \sum_{i=1}^{\infty} \frac{1}{\lambda_i} \langle A z_o, e_i \rangle^2 + 2 \cdot \sum_{i=1}^{\infty} \langle \tilde{z}_o(\theta), A^* e_i \rangle^2.$$

Gemäß Voraussetzung sind die Terme, die jeweils auf der rechten Seite der obigen Abschätzungen stehen, P-fast sicher endlich. Somit folgen (2.13) und (2.14) aus der exponentiellen Stabilität der Halbgruppe $(T_{\theta t})_{t \in [0,\infty)}$.

Schließlich verifizieren wir (2.15) und (2.16). Aufgrund der Hölderschen Ungleichung, (2.11), (2.12) und (2.17) reicht es aus zu zeigen, daß

$$(2.18)\quad \limsup_{T \to \infty} \frac{1}{T} \int_0^T \sum_{i=1}^{\infty} \frac{1}{\lambda_i} \langle A z(t, z_o, \theta), e_i \rangle^2 dt < \infty \text{ P-fast sicher}$$

bzw.

$$(2.19)\quad \limsup_{T \to \infty} \frac{1}{T} \int_0^T \sum_{i=1}^{\infty} \frac{1}{\lambda_i} \langle z(t, z_o, \theta), A^* e_i \rangle^2 dt < \infty \text{ P-fast sicher.}$$

Es gilt aber

$$\sum_{i=1}^{\infty} \frac{1}{\lambda_i} \langle Az(t,z_o,\theta), e_i \rangle^2$$

$$\leq 2 \cdot \sum_{i=1}^{\infty} \frac{1}{\lambda_i} \langle A(z(t,z_o,\theta)-\tilde{z}(t,\theta)), e_i \rangle^2 + 2 \cdot \sum_{i=1}^{\infty} \frac{1}{\lambda_i} \langle A\tilde{z}(t,\theta), e_i \rangle^2$$

und

$$\sum_{i=1}^{\infty} \frac{1}{\lambda_i} \langle z(t,z_o,\theta), A^* e_i \rangle^2$$

$$\leq 2 \cdot \sum_{i=1}^{\infty} \frac{1}{\lambda_i} \langle z(t,z_o,\theta)-\tilde{z}(t,\theta), A^* e_i \rangle^2 + 2 \cdot \sum_{i=1}^{\infty} \frac{1}{\lambda_i} \langle \tilde{z}(t,\theta), A^* e_i \rangle^2.$$

Also ist das Lemma angesichts der Beziehungen (2.11), (2.12), (2.13) und (2.14) bewiesen. □

Literaturverzeichnis:

[AK/TA, '81] M. Akahira/K.Takeuchi: "Asymptotic Efficiency of Statistical Estimators: Concepts and Higher Order Asymptotic Efficiency", Springer, New York, 1981.

[AN/KL/WO, '69] W.N. Anderson,Jr./G.B. Kleindorfer/P.R. Kleindorfer/ M.B. Woodroofe: "Consistent estimates of the parameters of a linear system", Ann. Math. Statist. 40, S. 2064-2075, 1969.

[AN, '71] T.W. Anderson: "The Statistical Analysis of Time Series", John Wiley, New York, 1971.

[AO, '67] Aoki: "Optimization of Stochastic Systems", Academic Press, New York, 1967.

[AR, '73] L. Arnold: "Stochastische Differentialgleichungen", R. Oldenbourg, München, 1973.

[AR/CU/KO, '80] L. Arnold/R. Curtain/P. Kotelenez: "Nonlinear Stochastic Evolution Equations in Hilbert space" Report No. 17, Forschungsschwerpunkt Dynamische Systeme, Universität Bremen, 1980.

[AS, '79] T. Assefi: "Stochastic processes and estimation theory with applications", John Wiley, New York, 1979.

[BA, '78] A. Bagchi: "Consistent estimates of parameters in continuous time systems" Memorandum Nr. 225, Technische Hogeschool Twente, Enschede, 1978.

[BA/BO, '81] A. Bagchi/V. Borkar: "Parameter identification in infinite dimensional linear systems", Memorandum Nr. 346, Technische Hogeschool Twente, Enschede, 1981.

[BAH, '67] R.R. Bahadur: "Rates of convergence of estimates and test statistics" Ann. Math. Statist. 38, S. 303-324, 1967.

[BA/RA, '80] I.V. Basawa/L.S. Prakasa Rao: "Statistical Inference for Stochastic Processes",
Academic Press, London, 1980.

[BE] D. Bell: "An infinite dimensional generalization of the Cameron-Martin-Girsanov Theorem",
Preprint.

[BI, '61] P. Billingsley: "Statistical Inference for Markov processes",
The University of Chicago Press, Chicago, 1961.

[BI, '68] P. Billingsley: "Convergence of Probability Measures",
John Wiley, New York, 1968.

[BO,JE, '70] G. Box/G. Jenkins: "Time Series Analysis, forecasting and control",
Holden-Day, San Francisco, 1970.

[BR, '74] A. Le Breton: "Parameter estimation in a linear stochastic differential equation",
Trans. of 7th Prague Conf. on Inf. Theory, Vol. A,
S. 353-366, 1974.

[BR, '76] A. Le Breton: "On continuous and discrete sampling for parameter estimation in diffusion type processes",
Math. Programming Study 5, S. 124-144, 1976.

[BU, '82] O.H. Bustos: "General M-estimates for contaminated pth-order autoregressive processes: consistency and asymptotic normality",
Z. Wahrscheinlichkeitstheorie verw. Gebiete 59, S. 491-504, 1982.

[CO, '81] D.R. Cox: "Statistical analysis of time series: Some recent developments",
Scand. J. Statist. 8, S. 93-115, 1981.

[CU/PR, '78] R. Curtain/A. Pritchard: "Infinite Dimensional Linear Systems Theory"
Springer, Berlin, 1978.

[DA, '67] Yu.L. Daletskii: "Infinite-dimensional elliptic operators and parabolic equations connected with them"
Russ. Math. Surv. 22, S. 1-53, 1967.

[DE/MA, '79] L. Denby/R.D. Martin: "Robust estimation of the first-order autoregressive parameter", J. Amer. Statist. Ass. 74, S. 140-146, 1979.

[DO, '53] J.L. Doob: "Stochastic Processes", John Wiley, New York, 1953.

[DU, '79] W. Dunsmuir: "A Central Limit Theorem for parameter estimation in stationary vector time series and its application to models for a signal observed with noise", Ann. Statist., S. 490-506, 1979.

[FR, '75] A. Friedman: "Stochastic Differential Equations and Applications", Bd. 1, Academic Press, New York, 1975.

[GÄ/ST, '77] P. Gänssler/W. Stute: "Wahrscheinlichkeitstheorie", Springer, Berlin, 1977.

[GI/SK, '79] I.I. Gihman/A.V. Skorohod: "The Theory of Stochastic Processes III", Springer, Berlin, 1979.

[GRI, '82] J.M. Griffith: "Identification of multivariate linear systems" Preprint des U.S. Department of Energy, 1982.

[GR, '65] L. Gross: "Abstract Wiener Space" Proc. Fifth Berkeley Sympos. Math. Statist. and Probability, Vol. 2, part 1, S. 31-42, 1965.

[GR/GR, '81] J. Grosspietsch/D. Graupe: "Convergence of least squares identifiers of time series with martingale difference and binary white Markov generating processes", Int. J. Systems Sci. 12, S. 119-125, 1981.

[HA, '70] E.J. Hannan: "Multiple Time Series", John Wiley, New York, 1970.

[HA/WI, '81] M. Hazewinkel/J.C. Willems (Hrsg.): "Stochastic Systems: the mathematics of filtering and identification and applications", Reidel, Dordrecht, 1981.

[HI/WA, '76] M. Hitsuda/H. Watanabe: "On stochastic integrals with respect to an infinite number of Brownian motions and its applications",
Proc. of Intern. Symp. SDE, Kyoto 1976, S. 57-74.

[IB/LI, '71] I.A. Ibragimov/Yu. V. Linnik: "Independent and stationary sequences of random variables",
Wolters-Noordhoff, Groningen, 1971.

[IK/WA, '81] N. Ikeda/S. Watanabe: "Stochastic Differential Equations and Diffusion Processes",
North-Holland, Amsterdam, 1981.

[JA, '70] A.H. Jazwinski: "Stochastic processes and filtering theory",
Academic Press, New York, 1970.

[KA/RA, '76] R.L. Kashyap/A. Ramachandra Rao: "Dynamic Stochastic Models from Empirical Data",
Academic Press, New York, 1976.

[KUN, '70] H. Kunita: "Stochastic integrals based on martingales taking values in Hilbert space",
Nagoya Math. J. 38, S. 41-52, 1970.

[KU, '75] H.H. Kuo: "Gaussian measures in Banach spaces",
Springer, Berlin, 1975.

[KUT, '75] Yu.A. Kutoyants: "On a hypotheses testing problem and asymptotic normality of stochastic integrals",
Theor. Prob. and Appl. 20, S. 376-384, 1975.

[LA, '79] V. Lanska: "Minimum contrast estimation in diffusion processes",
J. Appl. Prob. 16, S. 65-75, 1979.

[LI, '61] Yu. V. Linnik: "Die Methode der Kleinsten Quadrate in moderner Darstellung",
Deutscher Verlag der Wissenschaften, Berlin (DDR), 1961.

[LI/SH, '77] R.S. Liptser/A.N. Shiryayev: "Statistics of Random Processes I, General Theory",
Springer, New York, 1977.

[LI/SH, '78] R.S. Liptser/A.N. Shiryayev: "Statistics of Random Processes II, Applications"
Springer, New York, 1978.

[MAK, '80] S.Ya. Makhno: "Estimation of the drift parameter of the unobservable component of partially observable random processes",
Prob. of Inform. Transmission 16, S. 282-285, 1980.

[MA/WA, '43] H.B. Mann/A. Wald: "On the statistical treatment of linear stochastic difference equations",
Econometrica 11, S. 173-220, 1943.

[MA, '79] R.D. Martin: "Robust estimation for time series autoregressions", S. 147-176,
in R. Launer/G. Wilkinson (Hrsg.): 'Robustness in statistics',
Academic Press, New York, 1979.

[MA, '82] R.D. Martin: "The Cramér-Rao bound and robust M-estimates for autoregressions",
Biometrika 69, S. 437-442, 1982.

[ME/PE, '80] M. Metivier/J. Pellaumail: "Stochastic Integration",
Academic Press, New York, 1980.

[MÜ, '75] P.H. Müller (Hrsg.): "Lexikon der Stochastik",
Akademie-Verlag, Berlin, 1975.

[OU, '72] J. Ouvrard: "Processus à valeurs hilbertiennes et théorème de Girsanov",
C.R. Acad. Sci (Paris), t. 275, S. 1367-1370, 1972.

[OU, '73] J. Ouvrard: "Martingales locales et théorème de Girsanov dans les espaces de Hilbert réels séparables"
Ann. Inst. Henri Poincaré, Sect. B, Vol. IX, S. 351-368, 1973.

[PA, '74] M. Pagano: "Estimation of models of autoregressive signal plus white noise",
Ann. Statist. 2, S. 99-108, 1974.

[RA, '62] R. Ranga Rao: "Relations between weak and uniform convergence of measures with applications",
Ann. Math. Statist. 33, S. 659-680, 1962.

[SC, '66] L. Schmetterer: "Einführung in die Mathematische Statistik", Springer, Wien, 1966.

[SL, '76] M. Slemrod: "Asymptotic behaviour of C_0-semigroups as determined by the spectrum of the generator", Indiana Univ. J. 25, S. 783-792, 1976.

[TU, '80] J.K. Tugnait: "Identification and model approximation for continuous-time systems on finite parameter sets", IEEE Transactions on Automatic Control AC-25, S. 1202-1206, 1980

[UO, '79:1] K. Uosaki: "Sequential estimation of parameters of autoregressive time series", Technology Reports Osaka Univ. 29, S. 309-317, 1979.

[UO, '79:2] K. Uosaki: "Estimation of the parameters of the autoregressive time series using stochastic approximation", Systems Science 5, S. 173-185, 1979.

[WA, '69] M.T. Wasan: "Stochastic Approximation", Cambridge University Press, Cambridge, 1969.

[WI/NÖ, '70] H. Witting/G. Nölle: "Angewandte Mathematische Statistik", Teubner, Stuttgart, 1970.

Lebenslauf

14.8.1953	geboren in Gelsenkirchen als Sohn der Eheleute Heinz Loges und Maria geb. Baring
1960-1964	Besuch der kath. Volksschule Südfeldmark und Kirchschule in Wattenscheid
1964-1972	Besuch des städt. Jungengymnasiums Wattenscheid
Mai 1972	Abitur
1972-1979	Studium der Mathematik mit Nebenfach Physik an der Ruhr-Universität Bochum
April 1975	Vordiplom in Mathematik
April 1979	Diplom in Mathematik
April 1979 bis September 1979	Wissenschaftliche Hilfskraft an der Abteilung für Mathematik an der Ruhr-Universität Bochum
seit Oktober 1979	Verwalter der Stelle eines wissenschaftlichen Assistenten an der Abteilung für Mathematik an der Ruhr-Universität Bochum

Bochum, den 8.10.1982

DER DRUCKER - Buscheystraße 138 - 4630 Bochum - Tel. (02 34) 70 41 16